

Dangers of nationalism

The drive to acquire intellectual property from research contributes to the wealth of nations, but can also undermine science if carried to excess. Exaggerated claims threaten to undermine the funding and climate of basic research.

Many countries around the world see the acquisition of patents stemming from research as a matter of national security and economic strength. Government leaders and some scientists view information technology and biomedical research as key competitive fields, and they push, with reformist zeal, towards application. But the hype surrounding the relationship between national strength and science has taken on an ominous tone that threatens to harm both basic research and scientific internationalism.

Many countries today see themselves trailing the United States in an intellectual-property race, and are looking for ways to catch up. Moreover, there are moves directed against isolated scholars, who are cast as crusty conservatives pursuing projects to satisfy their own interest. Such researchers are criticized for failing to realize the potential applications, and for their lack of profit-oriented motivation.

For smaller or developing countries with less generous research budgets, the application imperative can be stultifying. Some Chinese researchers, for example, complain about their inability to get funding unless they have a good 'gimmick' to sell. They rightly worry that funding will increasingly depend on demonstrable application, while more technologically modest but scientifically notable projects are shelved.

Will good basic research end up being overlooked? One of the enthusiastic leaders of Japan's movement towards application warns: "We could very well look back in 50 years and say we've made a huge mistake. But now it is the right thing to do — if done with care."

Certainly, Japan is trying to create a moral imperative out of money-making. The push into technology licensing and the restructuring of government research organizations (see *Nature* **410**, 7; 2001) are geared towards making science more accountable to society and more profitable to researchers. They are also seen as preserving Japan's industrial strength — to some researchers in the vanguard, a lack of practical results means a waste of public money.

Still, much of this reform is just talk. Young revolutionary bureaucrats and technology-licensing promoters speak of radical change. Researchers write grant proposals in which they answer questions such as, "what benefit will your research have to society?". To a large extent, though, Japan continues to do a lot of good, basic research.

But even if bigger countries' basic research is protected from this latest phase of application-oriented reform, hype about intensifying industrial competition could restrict openness and thereby threaten another ambition that Japan, like many other Asian countries, pursues: respect for its researchers and their full participation within the international scientific community.

Former Prime Minister Yoshiro Mori's claim that Japan will overtake the United States in information technology within five years sounds like a harmless jest, but such hype creates a potentially explosive atmosphere. The recent 'bio-spy' incident, in which the United States charged Japanese researchers accused of stealing material for studying Alzheimer's disease with espionage, shows that the United States is increasingly ready to escalate tensions in such matters. Instead of being an incident involving the actions of a couple of researchers, accusation fell on Japan, sending shock waves through the Japanese research community. The Japanese press obligingly presented the episode as an almost inevitable outcome of an international intellectual-property race. Japanese researchers are understandably starting to wonder if they will ever be welcome in US laboratories. And as one US researcher in Japan put it, "all of the sudden I was frighteningly aware that I was a foreign scientist".

Countries who recognize that applied science can underpin their strength and prestige must also be prepared to support basic research, and to sustain the internationalism that is so essential to it. That requires a judiciously light touch in responding to foreign technological competition. ■

Europe's right approach to energy

The attitude of US governments and oil companies will be bad for their businesses.

Before last week's upheaval in US politics, big oil companies must have looked on their generous donations to George W. Bush's campaign as money well spent. Bush's energy plan, released the week before, called for more fossil-fuel power stations and the opening up of the Arctic National Wildlife Reserve for oil exploration.

The defection of Republican Senator James Jeffords looks likely to obstruct much of that (see page 510). But it is unlikely to change the attitudes of ExxonMobil and other US oil companies towards emissions reductions to ameliorate climate change. Their confident hostility is boosted by the lack of action by successive US administrations.

In Europe, the political climate is different. The European Union (EU), negotiating as a block at the Kyoto talks on combating climate change, appeared genuinely to want the agreement to work. Individual EU members have already begun legislating.

With further regulatory changes forecast, the European oil companies are moving to adjust their business models (see page 516). Sensing a future in which fossil fuels are not the dominant energy source, BP and Shell have invested in alternative sources such as solar and wind power. They have also pledged a 10% cut in greenhouse-gas emissions resulting from their operations, a target both say they will reach in the next few years.

This approach costs money now, but it will pay dividends as renewables become a bigger source of our energy. Fossil fuels cannot power the planet for ever, and the US companies should be trying to adapt. Without a strong lead from the US government, they look unlikely to change their ways. President Bush should recognize that his current approach will hurt more than just the environment. In the long term (and perhaps in the short term too, if consumer activists have their way), the US oil companies will also suffer. ■





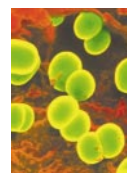
Power switch
Defecting Republican
hands Democrats
control of the Senate
p510



Future shock
Budget increases
offer the NIH a
double-edged sword
p511



Out of the frame
Eastern Europe fears
exclusion from EU
research programmes
p512



Fighting back
Antibiotic
resistance comes
under fresh scrutiny
p513

Data protection law threatens to derail UK epidemiology studies

David Adam, London

Epidemiological studies in Britain are being crippled by the overzealous application of legislation designed to protect personal data, claim leading researchers in the field. The epidemiologists warn that dozens of studies are under threat because doctors fear prosecution if they hand over information about their patients.

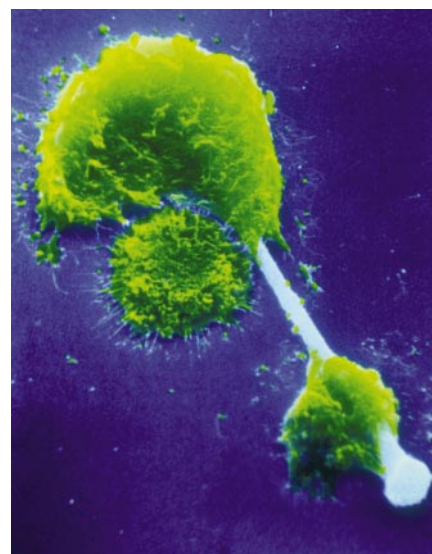
Britain is a world leader in epidemiology. But the 1998 Data Protection Act — which forbids personal data collected for one reason being used for anything else without informed consent — is threatening that position. “It’s blocking everything from follow-up studies of people who received medical treatment 20 years ago to finding out what industrial hazards may be causing cancer,” says Julian Peto, head of epidemiology at the Institute of Cancer Research near London.

Richard Doll of the Clinical Trial Service Unit at the University of Oxford, who in the 1950s established the link between smoking and lung cancer, calls the situation “potentially disastrous”; and adds that much of his research would never have been done in the present climate.

Peto says that his study of asbestos-related cancer, funded by the government through the Health and Safety Executive, is being obstructed because the Department of Health now refuses to tell him which doctors have eligible patients. “It’s ridiculous,” he says. “Talk about the left hand not knowing what the right hand is doing.”

Epidemiologists say matters have come to a head since September last year, when the General Medical Council (GMC), which regulates the medical profession, issued guidelines warning doctors that they risked prosecution if they passed on data without explicit consent. Until then, most doctors had assumed that supplying data to researchers was exempted by a ‘public interest’ clause in the law. “The GMC’s guidance has left the medical research community in turmoil,” says Michel Coleman, head of the cancer and public-health unit at the London School of Hygiene and Tropical Medicine. He wants the GMC to withdraw or modify its guidelines.

A spokesman for the GMC insists that it had no choice but to issue the guidelines in their present form. “After taking legal advice it became clear to us that doctors were acting



Gagged: data deprivation could halt studies on the health effects of asbestos fibres, seen here being attacked by cells of the immune system.

unlawfully in passing on confidential records without consent,” he says.

New provisions passed earlier this month offer a solution, suggests the GMC’s spokesman. A committee made up of patient groups and health professionals will discuss specific projects and advise the Secretary of State for Health, who has the power to waive the need for informed consent. But epidemiologists are unconvinced. “It remains unclear if the group can review hundreds of studies for approval without causing further delay or even the collapse of projects,” says Coleman.

Peto wants the government to issue a broad exemption for epidemiological research. “No harm was ever done in the past and the new law wasn’t even aimed at medical research,” he says.

Epidemiologists elsewhere are watching the situation. The UK law was designed to implement the 1995 directive on data protection from the European Union, and other member states have introduced or are preparing similar legislation. Concerns about the impact of data protection laws on epidemiology are also rising in Japan, South Africa and Canada.

Arson hampers conservation work

Rex Dalton, San Diego

A plant research facility at the University of Washington was severely damaged by fire last week, in what appears to be one in a series of destructive attacks by militant environmental activists.

The university’s Center for Urban Horticulture in Seattle, where ecologists, botanists and molecular biologists study biodiversity and plant genetics, suffered damage estimated at \$3 million in the incident on 21 May. Vehicles and buildings at a tree plantation in Clatskanie, Oregon, which works with researchers at the centre, were also damaged in a simultaneous attack.

The Federal Bureau of Investigation is investigating the fires, which an anonymous call claimed was the work of the Earth

Liberation Front (ELF). The letters ‘ELF’ were also daubed in graffiti at the Oregon plantation. The ELF has been linked to a series of attacks on industry, tourist and logging facilities in the Pacific Northwest.

The main target for the attack appears to be Toby Bradshaw, who heads a project studying genetically modified poplar trees. Bradshaw says that his research materials and data were backed up elsewhere.

But the work of other scientists, such as Sarah Reichard, who studies the restoration of endangered forest plants, has suffered. “These people are really ignorant, attacking scientists who spend their lives protecting the environment,” says Reichard, who is secretary of the Society for Conservation Biology.

Senate shift bodes well for 'green' science

**Tony Reichhardt, Washington
and Rex Dalton, San Diego**

Washington underwent an upheaval last week, when Senator James Jeffords of Vermont left the Republican Party and handed control of the US Senate to the Democrats. Science is not generally seen as a battleground in the shifting political landscape. But environmental policy will now move to centre stage, which may improve funding prospects for environmental research.

The shift will see new faces in the chairs of all key Senate committees. On the subcommittee that oversees funding of the National Science Foundation and NASA, Kit Bond (Republican, Missouri) is expected to give way to Barbara Mikulski (Democrat, Maryland). Leadership of the equivalent panel for the National Institutes of Health is expected to shift from Arlen Specter (Republican, Pennsylvania) to Tom Harkin (Democrat, Iowa).

On both committees, however, it should be business as usual. An aide to Bond, for instance, notes that he and Mikulski have "a close working relationship that has produced a reservoir of good will". And according to conventional US political wisdom, science often benefits when one party controls the White House and the other the Senate — spending on research is one subject on which opposing politicians can usually agree.



Wind of change: environmental science may find favour in the wake of James Jeffords' decision.

There could be greater implications in the shift in leadership of the subcommittee that oversees the Department of Energy's budget. Pete Domenici (Republican, New Mexico), whose state is home to the Los Alamos and Sandia National Laboratories, is expected to

be replaced by Harry Reid (Democrat, Nevada). Although Reid is a supporter of science, he has been critical of cost overruns at the National Ignition Facility, a huge laser project at the Lawrence Livermore National Laboratory in California.

Reid also opposes plans to develop his state's Yucca Mountain site as a repository for nuclear waste. And this is just one of a slew of environmental and energy policy issues on which the Democratic Senate majority will battle President George W. Bush. Plans to drill for oil in Alaska's Arctic National Wildlife Refuge, for instance, are now effectively dead, with the state's two pro-drilling Republican senators, Ted Stevens and Frank Murkowski, losing respective control of the Appropriations Committee and the Energy and Natural Resources Committee.

Jeffords, sitting as an independent, has reportedly been promised the chair of the Environment and Public Works committee. Such a move will please environmentalists, who respect his stance on green issues.

One result of all these changes is expected to be greater support in the Senate for environmental science and for research into alternative energy sources, both downplayed by Bush. If that occurs, future budgets may bring back smiles to the faces of researchers in these disciplines. ■

'Contrary' trade sanctions worry malaria researchers

Declan Butler

Fears of breaching US export controls have forced a centre that was set up as an international resource for malaria researchers to refuse requests for reagents from scientists in Sudan and Cuba.

The Malaria Research and Reference Reagent Resource Center (MR4) in Manassas, Virginia, was established in 1998 to supply information and reagents to malaria researchers across the globe. But scientists at the centre realized last month that they risk fines of \$11,000 for each reagent sent to Cuba or Sudan. Both countries are covered by a US trade embargo on the grounds that they pose

an "unusual and extraordinary threat" to US national security.

Sudan and Cuba are not the only nations subject to US trade embargoes, but they are most relevant to malaria research because of their geographical location. Although there are only a handful of malaria labs in each country, they are potentially important sites for clinical trials. Both countries have ongoing international collaborations.

David Walliker of the University of Edinburgh, who lobbied for the MR4's creation, says that had he realized the centre would be subject to US trade controls, he would have sought to have it funded by the World Health Organization, rather than the US government. "The embargo is contrary to the whole spirit and *raison d'être* of the MR4's establishment," he says.

The centre intends to ask the US Department of Commerce for an exemption from the embargoes. Research reagents do not at present qualify for exemptions granted to items such as food and medicine. Scientists fear that the MR4's future is at risk if the situation cannot be resolved. Alexandra Fairfield of the US National Institute of

Allergy and Infectious Diseases in Bethesda, Maryland, which funds the centre, fears that researchers may become reluctant to donate reagents to the centre — something the MR4 relies on to fulfil its functions.

Sudanese researchers have found their work hampered in other ways. In February, Moawia Mukhtar of the Institute of Endemic Diseases at the University of Khartoum had a paper on the parasitic disease onchocerciasis rejected by the US journal *Clinical Immunology*, which said: "Given the complexities of dealing with the export restrictions imposed by the United States government ban on trading with Sudan, and the severe penalties for running afoul of the law, we are not accepting contributions from Sudan."

Walliker wrote to the journal, protesting that the decision was contrary to the interests of international science. After taking legal advice, the journal decided this month that it could publish articles from Sudan, and has asked Mukhtar to resubmit. "It was so sad and humiliating for us," Mukhtar says. ■

♦ <http://www.malaria.mr4.org>



Disease dilemma: Sudanese researchers are falling victim to a US trade embargo.

NIH faces double trouble over budget rise

Matthew Davis, Washington

US biomedical science is experiencing unprecedented growth, with the National Institutes of Health (NIH) on course to have doubled its budget over five years by 2003. But there is rising concern that if, as is likely, budget increases then start to slow, too much of the agency's money will be locked into multi-year projects, squeezing funding for new grants.

Foremost among those expressing such worries is Senator Tom Harkin (Democrat, Iowa). Last week's power shift in the Senate, after a Republican senator's departure from his party handed control to the Democrats (see story opposite), is expected to see Harkin become chair of the Senate committee that oversees the NIH's budget. In committee hearings last week, Harkin said he was "really concerned" about the NIH's level of commitment to multi-year grants, if budgets tighten in 2004. "There will not be a lot of money left over for new grants," he warned.

The plan to double the NIH's annual budget to \$27 billion was drawn up partly to address complaints from biomedical researchers about the low success rate of grant applications. The number of NIH-funded research grants has since grown from 27,000 to 32,400, and their average value has risen by 36%.

Continuing this growth trend would please scientists. But with grants typically running for four years, it would leave the NIH with little spare cash for new projects if its budget growth slows.

Robert Rich of Emory University in Atlanta, president-elect of the Federation of American Societies for Experimental Biology, says multi-year obligations must be carefully managed if researchers applying for grants after 2004 are not to be short-changed. But even with smart planning, he claims, problems will only be avoided if the NIH's budget continues to increase at



Headache at headquarters: is the NIH locking too much of its money into multi-year research grants?

around 7% per year. And that may not be realistic, given the US administration's moves to cut taxes, and concerns about a general economic slowdown.

Last week, NIH acting director Ruth Kirschstein told Harkin that the agency plans to set up a committee of institute directors to consider future spending for various projected budget scenarios. But congressional aides say privately that such planning may come too late. And they are concerned that NIH officials are still assuming they will get a succession of multi-billion-dollar increases after 2003.

Francis Collins, who heads the National Human Genome Research Institute in Bethesda, Maryland, agrees that there could be a

problem. His institute's budget is heavily committed to projects spanning several years, and the effect on new grants, should NIH budget growth stall, "would be gruesome", he claims. Collins says one approach would be to invest more in facilities, equipment and an array of short-term projects this year and next. But that would mean fewer research grants.

This would be unpopular with researchers who want to benefit from the NIH's current growth. Making the tough decisions that may be necessary to plan for the future would be easier if the agency had a permanent director. But so far the US administration has made no moves to propose a candidate. ■

Dust settles on defamation case

Rex Dalton, San Diego

A long-running battle over claims that a scientific paper defamed the inventor of a radiocarbon-dating technique has finally been settled.

Under the agreement, the state of Arizona, which funds the universities employing both parties to the dispute, will pay Ronald Dorn of Arizona State University in Tempe \$50,000 to cover legal costs.

Dorn developed his technique to date scrapings of the natural 'varnish' that forms on desert rocks. But a paper in *Science* (280, 2132–2139; 1998), by authors including Warren Beck, Douglas Donahue, Tim Jull and George Burr of the University of Arizona, Ekkehart Malotki of Northern Arizona University, and Wallace Broecker of Columbia University in New York, reported "ambiguities" in samples obtained from Dorn's lab. The paper concluded that

the samples were contaminated with fragments of charcoal and coal — which would have affected the dating measurements.

Dorn became the subject of a misconduct investigation, but was cleared by his university of manipulating his samples in September 1999. In the meantime, he sued the paper's authors for defamation. Broecker settled with Dorn last year (see *Nature* 408, 506; 2000), but Dorn's dispute with the Arizona authors was, until last week's settlement, heading for an October jury trial.

The settlement means that neither side will risk running up massive legal bills. "Dorn and the authors agree that science is developed through discussion and debate in the literature," says a statement issued by the University of Arizona. Dorn declined to comment. ■



Eastern Europe decries EU research proposal

Quirin Schiermeier, Linköping, Sweden

Proposed changes to the structure of the research programme run by the European Union (EU) could torpedo the efforts of central and eastern European countries to increase their participation, scientists and policy-makers warned last week.

Most of the EU's research activities are organized into five-year 'Framework' programmes. Until now, these have supported large numbers of relatively small projects, creating networks that link scientists from several EU member states — and nations that have 'association' agreements with Brussels.

But the European Commission plans to shake up the 17.5-billion-euro (US\$15 billion) sixth Framework, scheduled to start in 2003. Research commissioner Philippe Busquin wants to focus on larger, integrated projects (see *Nature* **410**, 4; 2001). And he intends to delegate the management of each project to one of the research centres involved.

Researchers from the 11 'candidate' nations waiting to join the EU believe these changes will favour countries that already have a strong science base. Their complaints spilled over last week in Linköping, Sweden, at a meeting held to discuss the role of the candidate countries in EU research.

Bulgaria, the Czech Republic, Cyprus, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia are all scheduled to join the EU in the next few years — and scientists from these countries are already eligible for funding under the fifth Framework. Each candidate country pays an entrance fee calculated according to its gross

domestic product and spending on research.

Even under the existing system, researchers from candidate countries have found it difficult to compete for funding. Poland estimates that its scientists have recouped only 60 million euros of its entrance fee of 160 million euros.

"We expected a mountain of gold in Brussels just waiting for us to go and fetch it," says Andrzej Siemaszko, who heads the government office in Warsaw that helps Polish scientists write EU project proposals. But given low success rates, "it is getting increasingly difficult to convince people to spend three months on preparing an application", he says.

The new Framework proposals are expected to widen the gap between the candidate countries' investment and their returns. If this happens, some observers fear that these countries may withdraw from their association agreements — dealing a serious blow to prospects for greater European cooperation in research.

"It took us more than two years to acquire the skills needed to write a good proposal," Uros Stanic, an electronics engineer at the Josef Stefan Institute in Ljubljana, told last week's meeting. "The switch to a completely new terminology and new funding instruments will force us to start from scratch."

Last month, Poland, the Czech Republic, Slovakia, Slovenia and Hungary called on the European Commission to involve more representatives of candidate countries in preparing the sixth Framework, and to reconsider the balance between small and large-scale projects.

EU officials deny that the integrated projects will exclude scientists from the candidate countries. "They do not necessarily require lots of expensive infrastructures and equipment," says Hendrik Tent, deputy director of the commission's research directorate. "They can also be clusters of smaller projects designed to meet one coherent objective."

But the commission is trying to address the candidate countries' concerns. It plans to spend 40 million euros on moves to encourage contacts between their scientists and those in EU member states, and is also planning a series of 'information days' to explain funding procedures.

♦ <http://europa.eu.int/comm/research/nfp.html>



Researchers at Budapest's Eötvös University are among those who fear they may be excluded.

Astronomers find fast-food plans hard to swallow

Sally Goodman, Paris

Astronomers at France's Pic du Midi observatory in the Pyrenees claim that plans to open a fast-food restaurant at the site will

have disastrous effects on their research.

François Colas and Jean Lecacheux of the Paris Observatory, who use a 1-metre telescope at Pic du Midi to study comets and planetary systems, have launched a petition to stop the development. They say that hot air and greasy fumes will damage the telescope's mirror — which is situated just 15 metres above the planned restaurant.

Ironically, the observatory was saved from closure in 1998 through a scheme to develop the site for tourism, after the French government threatened to withdraw funding from telescopes less than 2 metres in diameter. Until now, astronomers have stoically accepted the changes. But the fast-food restaurant plan, hatched by regional and local authorities, is "the last straw", says Colas.

The situation is symptomatic of the plight of observatories operating small

telescopes, whose futures are threatened by the proliferation of telescopes with mirrors larger than 8 metres across. These huge facilities consume the lion's share of most national astronomy budgets (see *Nature* **408**, 12–15; 2000).

But there is still valuable science to be done with small telescopes, argues Gerry Gilmore of the University of Cambridge, who chairs OPTICON, the Optical Infrared Coordination Network, an initiative funded by the European Union.

"It is often not cost-effective to do specialized experiments on the very large telescopes," Gilmore says. Projects that need repeated observations of objects, for instance, usually cannot get time on big telescopes. Smaller telescopes also play a vital role in training, he adds.

♦ <http://www.obs-mip.fr/omp>

♦ http://www.bdl.fr/s2p/tourisme/tourisme_engl.html



Cold shoulder: astronomers at Pic du Midi say greasy fumes could ruin their research.

Gene tests lift lid on drug-resistance puzzle

Jonathan Knight, Orlando

Is the use of antibiotics in farm animals causing the spread of antibiotic-resistance genes to human pathogens? The jury is still out, say experts who gathered last week at the annual meeting of the American Society for Microbiology in Orlando, Florida. But this long-standing controversy could soon be resolved, thanks to the development of molecular tools to help recognize resistance genes that originated on the farm.

Infections that resist treatment with antibiotics are a growing problem. Most researchers agree that over-prescription of antibiotics by doctors is the main cause.

But antibiotics are also routinely added to the feed of farm animals as minor infections can stunt their growth. This is promoting the emergence of antibiotic resistance among bacteria that inhabit or infect livestock. And such bacteria have been shown to cause some cases of human food poisoning.

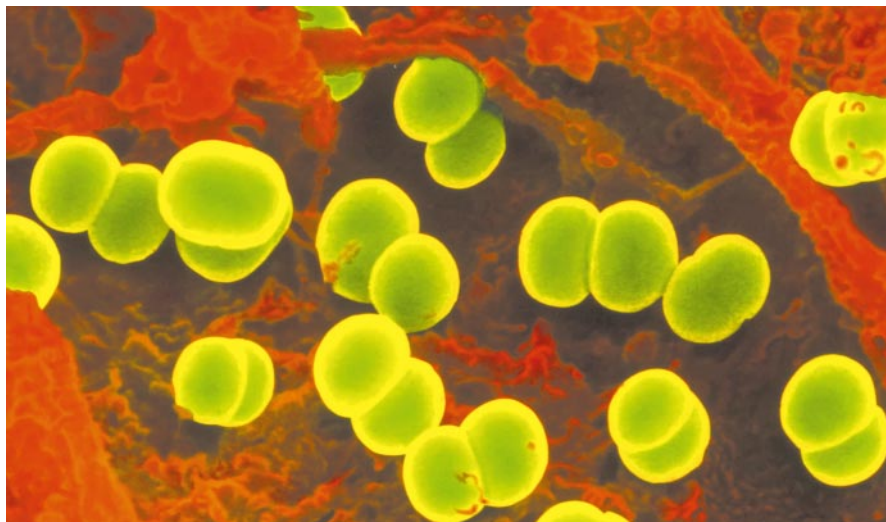
More worrying, however, is the possibility that farmyard antibiotic-resistance genes could be transferred to the bacteria that inhabit our bodies, creating an insidious reservoir of drug resistance. But it has proved difficult to demonstrate whether or not such gene transfer is occurring.

"It's hard to get agreement on the extent of the problem," says Walter Hill, an official with the US Department of Agriculture's Food Safety and Inspection Service.

Among those investigating the connection between farm antibiotics and human health is David Wagner, an animal scientist at the Food and Drug Administration's Center for Veterinary Medicine in Laurel, Maryland. He collected microorganisms from beef and poultry at local supermarkets, and found that two-thirds of one species of gut bacterium, *Enterococcus faecium*, sampled from chicken, were resistant to Synercid. This antibiotic has only been on the market for two years, and is used as a last resort to treat infections that resist the more commonly used vancomycin.

Although Synercid is not used in agriculture, its close cousin virginiamycin has been given to cattle, pigs and poultry for more than 20 years in the United States and Europe. Genes for virginiamycin resistance also confer resistance to Synercid, and the possibility that they can be transferred to human pathogens is a serious concern. "This is not good news," says Wagner of his findings.

Synercid treatment currently fails in only a tiny proportion of cases. But Stuart Levy of Tufts University in Boston, whose work on antibiotic resistance helped to open the field, advises caution. If resistance genes do transfer from livestock bacteria to those infecting humans, they could spread rapidly if Synercid becomes more widely used.



Meaty issue: are farming practices to blame for resistance in human pathogens such as *Enterococcus*?

Other researchers are now using molecular techniques that can rapidly recognize specific mutations involved in antibiotic resistance. Roustam Aminov of the University of Illinois at Urbana-Champaign has developed tests, based on the polymerase chain reaction, that can identify eight different classes of gene that confer resistance to tetracycline. Bacteria from cows and pigs carry different characteristic combinations of resistance genes, and Aminov has used his tests to track the transfer of tetracycline-resistance genes from pig farms in Illinois to soil bacteria.

Mark Maiden, a geneticist at the Forsyth Institute in Boston, Massachusetts, has used similar tests and found matching tetracycline-resistance genes in human mouth bacteria and in bacteria from animal intestines. "You've got identical sequences turning up in

quite different species," he says. "That's highly symptomatic of horizontal gene transfer."

Although oral bacteria cause nothing worse than dental cavities, Maiden says they could serve as a reservoir of tetracycline-resistance genes that might transfer to opportunistic bacteria that cause pneumonia or postoperative infections, making these more difficult to treat.

Maiden says he cannot prove that the genes did transfer from the animals' bacteria to the human oral bacteria—in theory, it could have been the other way around. But microbiologists are optimistic that the application of such molecular methods by more researchers will soon reveal the extent to which the transfer of farmyard resistance genes to human pathogens is occurring. "We've needed this for many years," says Levy. ■

Bioinformatics to be nurtured online

Declan Butler

Attempts to address the worldwide shortage of bioinformaticists have received a boost with the announcement that six academic centres will provide free online accredited bioinformatics courses.

The pre-release website for the Global Genomics and Bioinformatics Unified Learning Environment (GLOBULE) is already available, and features a series of video lectures from some of the biggest names in bioinformatics. They include Russ Altman of Stanford University in California and Winston Hide, director of the South African National Bioinformatics Institute at the University of the Western Cape.

Pitched at third-year undergraduate level and above, the site will allow students

to interact with teachers and other students around the world by e-mail and through chat rooms, says Tony Weiss, director of the Molecular Biotechnology Program at the University of Sydney in Australia. The finished version will offer a range of modules that will be credited by members of the consortium.

The programme could make a dramatic difference to the growth of genomics in developing countries, says Hide.

The founding members of the 'S-Star' consortium are Stanford University, Sweden's Uppsala University and Karolinska Institute, the National University of Singapore, the University of Sydney and the University of the Western Cape. ■

▶ <http://www.s-star.org>

Initiative promises stricter controls on clinical trials

Washington A new body set up to run a voluntary accreditation procedure for US research involving human subjects was unveiled last week.

The Association for Accreditation of Human Research Protection Programs (AAHRPP) will provide accreditation based on site visits and review by a board of directors. The AAHRPP is the brainchild of Sanford Chodosh, president and founder of the non-profit organization Public Responsibility in Medicine and Research. He promises that its standards will go beyond the existing federal regulations.

The initiative is the latest in a series of proposals for reforming the existing patchwork of regulatory procedures (see *Nature* **410**, 1017; 2001). The system was heavily criticized in 1999 after the death of a volunteer in a gene-therapy trial.

The do-nothing pills that really do nothing

Copenhagen The placebo effect is all in the mind — although not in the way most people think. After reviewing thousands of medical



The placebo effect — an imaginary friend?

papers from the past 50 years, Danish researchers say that the dummy drugs given to control patients do not help them to recover. Asbjorn Hrobjartsson of the University of Copenhagen and Peter Gotzsche of the Nordic Cochrane Centre, also in Copenhagen, identified 114 studies in which researchers investigating potential treatments had used two control groups — one that took placebos, and another that took nothing. Across these studies Hrobjartsson and Gotzsche could find no differences between the two types of control group (*New Engl. J. Med.* 394, 1594–1602; 2001).

Shock as labs miss anthrax samples

Orlando Four clinical laboratories in New Mexico failed to spot samples of anthrax sent to them as part of a test by the state's health department, last week's meeting of the American Society for Microbiology in Orlando was told.

Debra Horensky and Linda Nims of the Scientific Laboratory Division of the New Mexico Department of Health in Albuquerque submitted cultures of a weakened version of the anthrax bacterium to four local labs for analysis. All four labs identified the correct genus, *Bacillus*, but none recognized it as the deadly *B. anthracis*. Most species of *Bacillus* are harmless. Only one lab sent the sample on to the health department for further analysis, and then only after nine days.

Horensky has since trained the four labs in the use of a test for anthrax, which should be performed on the day the sample comes in.

Boost in German funds for prion disease studies

Munich The German government has announced a tenfold increase in funding for research into transmissible spongiform encephalopathies (TSE), which include mad cow disease, scrapie and the human Creutzfeldt–Jakob disease.

The new money — DM27 million (US\$11.8 million) per year — will be used to fund a programme aimed at developing high-sensitivity tests for TSEs in living animals and therapies for treating human TSEs. The research, which includes a plan to deliberately infect and monitor around 50

cattle, to be based at the Federal Research Centre for Virus Diseases of Animals on the island of Riems in the Baltic Sea.

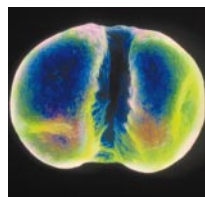
Research on the risks of humans contracting TSEs from infected animals will continue and will be extended from food studies to include the safety of products made using animal tissues, such as cosmetics.

Japan plans pollen purge to fight allergies

Tokyo The city government of Tokyo is planning to give the city's cedar trees shots of a plant-growth regulator in a bid to reduce the amount of pollen in the air.

An estimated one in every five Tokyo residents suffers allergic reactions to the pollen produced by the cedar trees, which cover 20,000 hectares of the city.

Experiments with young trees have shown that an injection of maleic hydrazide into the trunk produces a 96% reduction in the number of male flowers, the source of the pollen. The Tokyo Metropolitan Agricultural Experiment Station



Pollen grains make life a misery for many.

plans to extend its research to fully grown trees this summer. If successful, a city-wide programme of injections could begin in 2003.

Environmental groups have expressed concern over the possible long-term effects of the treatment. A more environmentally friendly, but longer-term project would replace the pollen-heavy cedars with a different variety of cedar that produces less pollen.

It's time to talk, but the name's a bit of a mouthful

Munich Yet another talking shop? No, say members of the **European Intergovernmental Research Organizations Forum (EIROFORUM)**, launched last week by the heads of seven leading European research centres, including the **European Space Agency and the European Molecular Biology Laboratory (EMBL)**.

"It's not an astonishing idea — it's the end of an abnormal era where large research centres didn't talk to each other," says Dirk Dubbers, head of the Institut Laue-Langevin, a neutron-source facility in Grenoble, France. Heads of the research centres have been meeting informally for over a year.

Discussions have already led to the creation of the Partnership for Structural

Biology, charged with overseeing the creation of a new neutron beam at the Institut Laue-Langevin and a new X-ray beam at the European Synchrotron Radiation Facility in Grenoble, both for the use of EMBL biologists studying molecular structure.

Consensus over the correct pronunciation of EIROFORUM has yet to be reached, however.

Flatulence levy is gone with the wind

Wellington Plans to combat climate change by raising a 'flatulence tax' on cows and sheep were dropped last week by the New Zealand government. Instead, it called for more research to improve the animals' digestion.

A ministerial group on climate change has targeted the methane released by New Zealand's 9 million cattle and 46 million sheep in a bid to reduce greenhouse-gas emissions to 1990 levels. But proposals for a tax of NZ\$15 (US\$6.4) per sheep and NZ\$60 per cow were greeted with fury by farmers, who argued that it would force many farms into bankruptcy.

Pete Hodgson, the science minister, now says that research into bovine digestion and pasture composition is the best way to cut emissions from the animals.

A change of climate for big oil

As European environmentalists launch a boycott of US oil firms, other energy companies are winning praise for their efforts to tackle climate change. Mark Schrope examines the oil giants' divergent strategies.

It is rare indeed to find an oil executive who has the respect of environmentalists — but in May 1997, John Browne, group chief executive of the London-based oil company BP, became just that. Speaking at Stanford University in California, Browne was the first senior executive of an oil company to acknowledge that carbon dioxide and other greenhouse gases probably caused global warming, and to recommend that action be taken.

It was a brave statement. Both BP's manufacturing processes and its products generate CO₂, and by explicitly linking emissions of the gas to global warming, Browne was associating his company with perhaps the biggest environmental threat to our planet. "A lot of people viewed BP as a traitor to the industry," says Eileen Claussen, director of the Pew Center on Global Climate Change in Washington, a non-profit organization that promotes market-based approaches to tackling climate change.

Four years on, climate-change initiatives taken by BP and the Anglo-Dutch oil company Royal Dutch/Shell have split the industry and earned the two firms praise from environmental lobbyists. But US companies, most notably ExxonMobil (known as Esso outside the United States), face a summer of protests in Europe. Environmental groups accuse these companies of influencing President George W. Bush's decision to reject the Kyoto Protocol — the international treaty that aims to limit emissions of greenhouse gases — and hope to mobilize a consumer backlash against the firms.

Despite the differences in environmental policy, profits are the ultimate measure of success for all the companies. Each oil giant is facing a dilemma — should it gamble on investments that may prepare it for the future, when concerns about global warming have severely restricted the burning of fossil fuels, or reap maximum short-term rewards while the oil business lasts in its present form?

Alternative energy sources provide only a small fraction of the world's energy.



Future investment: in 1997, BP's chief executive John Browne (top) acknowledged the role of greenhouse gases in global warming. Since then, the company has invested in renewable energy such as this solar-powered lighting rig (top right) used at the Sydney Olympics.

All the big oil companies now accept that greenhouse-gas emissions pose a potential threat. All have invested in technologies to reduce such emissions, from techniques for capturing CO₂ before it is released into the atmosphere, to alternative methods for producing energy, such as solar power.

But behind the companies' environmental rhetoric, their approaches are very different. Only BP and Shell have made firm commitments to reducing their emissions of greenhouse gases and backed them up with targets and deadlines. They are also making the heaviest investments in alternative sources of energy.

Central to BP's efforts is a commitment to knock 10% off its 1990 level of greenhouse-gas emissions by 2010. Working with the New York-based organization Environmental Defense, the company has set rigid yearly emissions targets for each of its business units, from drilling operations to chemical plants.

Key to the scheme is the idea that the reductions do not have to be spread evenly around the business. Instead, cuts in emissions are treated as commodities that can be bought and sold. Under the plan, CO₂ emissions are priced at \$8 per tonne. If a par-

ticular unit finds that the cost of cutting its emissions is less than this, it can aim to beat the 10% target and sell any excess 'carbon credits' on an internal market hosted on the company's intranet. Units that find the cost of cutting their CO₂ emissions is greater than \$8 per tonne can choose to buy carbon credits, which count towards their 10% target. If it suits them, units can buy their way out of making any cuts. Credit is also given for lowering emissions of other greenhouse gases, based on their warming potential relative to CO₂.

Jeff Morgheim, BP's climate-change manager, handed out the reduction targets across all units uniformly, regardless of how hard reductions would be for a given group. The process did not endear him to the group leaders. "That's why I wear shades and a trench coat," he jokes.

But the group leaders have now come round, Morgheim argues. The company is already halfway towards meeting its targets, and estimates that it should reach its goal ahead of schedule, probably by 2003. Perhaps more importantly, the overall cost of the programme has been kept down by units finding reductions that stem from efficiency improvements. Morgheim points out that the second phase will be more expensive, as units made the easiest and cheapest savings first, but he is confident that the scheme will succeed.

Trading places

The element of profit in the system has also sparked some innovative ideas. Morgheim cites a recent call from a refinery manager who was considering using the heat generated by his refinery to heat water for a nearby community, and wanted to know if he would receive emission credits for the energy saved by the community. "When I hung up the phone I knew we had made the right decision," Morgheim says.

Shell, also in conjunction with Environmental Defense, has set itself an even more ambitious target of a 10% cut in 1990 emissions levels by 2002. Again, it is using a system of internal emissions trading, and looks set to meet its target. According to the Pew Center, BP's cut will take longer to achieve as more time is needed to incorporate some of its newly acquired subsidiaries into the programme.

Both schemes will be watched closely by negotiators seeking to salvage the Kyoto Protocol. If the accord is ever ratified, a similar system for 'emissions trading' will have to be implemented on an international scale. Arguments over the details of such a scheme contributed to the breakdown of the last set of international climate talks. "One of the reasons why we set up an internal emissions trading system ourselves was to see whether those systems can work in practice," says Robert Kleiburg, a climate-change analyst with Shell.

Shell has also developed an impressive



Fuel for debate: Greenpeace's Steve Sawyer (right) welcomes the strategies employed by BP and Shell but backs the boycott on the anti-Kyoto US companies such as Exxon/Esso.

roster of alternative energy projects through a US\$500 million investment in its subsidiary Shell Renewables. Last year, together with a consortium of electricity generating and engineering companies, Shell Renewables opened a two-turbine wind farm off the coast of northeast England. The 4 megawatt (MW) farm is supplying electricity to 3,000 households. The company is now considering plans for a bigger project in the Irish Sea.

Working with Swedish energy company Sala-Heby Energi, Shell Renewables is also developing environmentally friendly combined heat and power (CHP) plants. Usually powered by gas, CHP plants are highly efficient and can provide both heat and electricity for anything up to a small town. Last August, Shell Renewables unveiled a CHP plant powered by wood chips left over from timber businesses which can provide 10 MW of electricity and 22 MW of heat. Hydrogen-powered fuel cells, batteries that many believe will run the electrical cars of the future, have also attracted investment from both BP and Shell. Other smaller oil companies are investigating carbon sequestration (see "The North Sea bubble", overleaf).

Despite these projects, alternative energy sources provide only a small fraction of the world's total energy needs. BP Solar, now the largest solar power company in the world, sold 40 MW of new capacity last year. This is a reasonable slice of the 250-MW global total, but with some individual fossil-fuel power stations generating thousands of megawatts, it is clear solar power still has some way to go.

Other oil companies seem to have turned their backs on renewables. ExxonMobil, for example, invested US\$500 million in alternative energy before deciding that the practical and economic challenges of turning renewables into profitable energy sources



were too great to warrant further funding.

Alternative sources may have a long way to go, but environmentalists are giving cautious praise to the efforts of BP and Shell. "I think they are way ahead of the rest," says Claussen. Steve Sawyer, an Amsterdam-based climate campaigner with Greenpeace International, wishes both companies would pursue alternative strategies more aggressively, but he backs their current efforts. "It's more than just greenwash," he says.

Such approval would have been out of the question in the early 1990s. As awareness of climate-change issues was starting to grow, many oil companies, including Shell and BP, funded lobbying groups such as the Global Climate Coalition (GCC). The GCC disputed the emerging scientific evidence for global warming and portrayed attempts to limit emissions as harmful to the world economy. BP and Shell have since withdrawn from the GCC, saying that they are now convinced that the science points to a warming world.

But ExxonMobil continues to speak out against the Kyoto Protocol and questions the degree to which the Intergovernmental Panel on Climate Change (IPCC), the group of scientists that advises governments on climate-change issues, represents a true consensus.

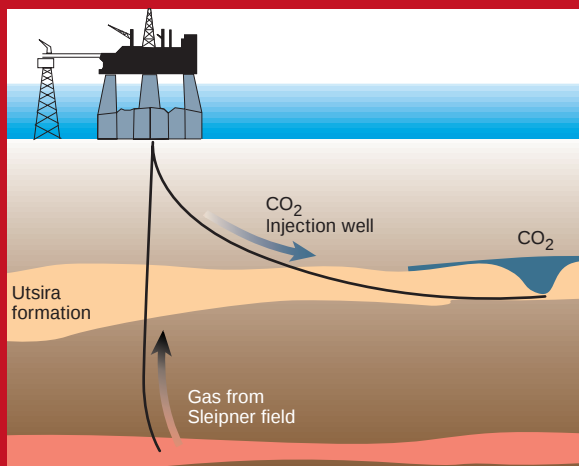
The North Sea bubble

Fossil fuels are likely to provide most of our power for the foreseeable future, but that does not mean that all the CO₂ produced has to go into the atmosphere. Norwegian company Statoil is pioneering attempts to store the gas elsewhere — deep below the seafloor. As Statoil's underground gas bubble grows, so do hopes that the technique could deal with more of the troublesome greenhouse gas.

The CO₂ is an impurity in the natural gas that Statoil extracts from the North Sea Sleipner gas fields. It is normally separated and released into the atmosphere, but since 1996 the company has been pumping the gas into a layer of sandstone around 1 kilometre below the seabed. Known as the Utsira formation, the layer traps the gas in a gigantic bubble which now contains some four million tonnes of CO₂.

Once separated from the natural gas, the CO₂ is compressed before being pumped into the reservoir, where the high pressure keeps it in a dense 'supercritical' state — a hybrid of gas and liquid phases. This limits the diffusion of CO₂ through the sandstone. A layer of shale, which is impermeable to the CO₂, sits on top of the sandstone, effectively sealing the reservoir.

Andy Chadwick of the British Geological Survey is part of a team of European scientists



responsible for monitoring the project. The team's 1999 seismic survey showed that the CO₂ is trapped within the reservoir and will probably stay there indefinitely. Chadwick is impressed with the project so far, but warns that the CO₂ may start to compress, rather than dissipate through the reservoir, making it progressively harder and more expensive to inject more gas.

But the Sleipner project's success does not necessarily imply that similar schemes could be applied to CO₂ from power plants. At Sleipner, the CO₂ would have been extracted from natural gas anyway, regardless of any storage scheme. Separating the CO₂ produced by power plants would require new investment, increasing the price of electricity. Additional infrastructure would

also be needed to transport the gas to the reservoir. But the potential for storage is huge: just 1% of the reservoir Statoil is using, says Chadwick, could hold three years' emissions from all of Europe's power stations.

For Statoil, the Sleipner project has actually saved the company money. Norway taxes offshore carbon emissions to the tune of \$38 per tonne. Before January 2000, the price was \$50 per tonne. By storing one million tonnes of CO₂ undersea every year and avoiding this tax, Statoil recouped its \$80 million investment within two years. The company plans to run the project for 20 years, storing around one million tonnes of CO₂ every year — equivalent to 3% of Norway's current total annual emissions.

David Adam

Steve Cochran, a spokesman for Environmental Defense, agrees. "Exxon has been extraordinarily difficult," he says. "It has funded efforts such as the GCC, which has worked very hard to undermine legitimate scientific understanding."

Greenpeace and other environmental groups have reacted by launching a European boycott of ExxonMobil and other American oil companies such as Texaco and Chevron. Sawyer says the goal is to provide encouragement for companies that do support Kyoto. In the long term, they hope to convince the United States to reconsider its decision.

If the boycott works, it may force ExxonMobil to factor consumer demand for environmentally friendly products into its long-term plans. Balancing environmental demands from the public and governments with the need to maintain profits will be a key challenge in the future for the oil industry.

Returning to Stanford earlier this year to give an update on BP's work on climate change, Browne described how the job of providing energy can feel like a trade-off between economic growth and a healthy environment. "I believe there is a huge commercial prize for those who can offer better choices which transcend the trade-off," he said.

The divergent climate-change policies of the oil giants are, in effect, speculations that reflect different views about that prize. Shell and BP are trying to position themselves as the main providers of the energy sources that may succeed fossil fuels. "They want to be there first," observes Claussen. ExxonMobil, on the other hand, appears to have decided that it can make more money by continuing under the present system for as long as possible — and catching up on alternative energy technologies at a later date.

Gordon Edge, an analyst with FT Energy in London, says ExxonMobil's decision will help profits in the short term, but will cause problems as renewables become more important. "Adjusting to renewables involves a major culture change. Exxon could take a decade to catch up," says Edge.

The long-term payoffs of all the companies' strategies are difficult to judge. But what is clear is that the outcome will affect us all: in gambling on their business futures, the oil giants are staking the Earth's climate.

Mark Schrope is a freelance writer in Melbourne, Florida.



Pipe dreams: US companies seem reluctant to invest in renewable sources to replace fossil fuels.

The latest IPCC report, issued earlier this year, stated that global temperatures were rising faster than previously thought and that there was strong evidence that greenhouse gases were the cause (see *Nature* 409, 445; 2001). But Frank Sprow, vice-president of safety, health and the environment at ExxonMobil, says that most of the evidence linking greenhouse gases to global warming comes from climate models, rather than empirical data. "I think the layman and even some scientists don't understand that the question of attribution has to rely on models," Sprow says. He also claims that the report's executive summary played down substantial uncertainties within the document.

Joyce Penner, an atmospheric modeller at the University of Michigan in Ann Arbor, denies this. "The summary for policy-makers does tell the story," she says. Penner was a lead author on the IPCC report chapter dealing with atmospheric aerosols, one of

the largest sources of uncertainty in understanding climate change. "If you choose to ignore some of the things said there then you will get a skewed view," she says.

Pouring oil on troubled waters

Sprow does not deny that the build up of greenhouse gases poses long-term risks, but he argues that the Kyoto Protocol is not the way to deal with them. Instead, ExxonMobil advocates a three-tiered voluntary response to climate change. The initial focus is on energy conservation, followed by advances in technologies that reduce CO₂ emissions, and finally breakthroughs in areas such as carbon sequestration, which aims to lock emitted CO₂ away so that it cannot enhance the greenhouse effect.

Sawyer says such a strategy is unacceptable: "All it means is delay and taking actions which are of no conceivable cost to them at some undefined point in the future."

Toronto's science jewel

A small Canadian institute is producing a disproportionate number of highly cited biology papers. Trisha Gura visited the Amgen Institute, to find out what its members are doing right.

Each spring, the newsletter *Science Watch* unveils a list that must have some scientists green with envy. Compiled by ISI, formerly the Institute for Scientific Information in Philadelphia, the list documents those researchers who have produced the most 'hot' papers over the preceding two years — those judged by a bespoke computer algorithm to be cited markedly more in the scientific literature than comparable papers.

Thirteen individuals who produced five or more such papers in 1999–2000 headed this year's roll of honour. Three of them are based at the Amgen Institute in Toronto. If citation analysis is a valid measure of scientific quality, then this small research centre, which specializes in disrupting mouse genes, must be one of the most distilled concentrations of excellence on the planet. It has only six faculty members, and just 50 scientists in total, including graduate students. So what explains its success?

In part, the answer lies in annual funding of some \$8 million from Amgen, the biotech giant of Thousand Oaks, California. But researchers familiar with the institute say that another defining factor is the personality and scientific style of its director, Tak Wah Mak.

Seeds of success

The story began in the early 1980s, when Mak was working at the Ontario Cancer Institute (OCI) in Toronto. Canadian funding for biomedical research was far from generous, and Mak had to scrimp and save to pursue his work on leukaemias triggered by retroviruses. In an attempt to garner more funding, he began to work on the immune system's T cells, which play an major role in destroying virus-infected cells.

The big break came in 1984, when researchers led by Mak became one of the two teams that won the race to clone the gene for the main receptor that enables T cells to lock onto their targets¹. The paper caught the eye of officials at Amgen, and the soft-spoken but shrewd Mak used the access this gave him to the company's executives to discuss a risky academic-industrial partnership.

Mak's idea was to create a research institute that would advance fundamental research while giving Amgen information about potential drug targets — the company would own any patents arising from the work, even if it was supported in part by outside grants.

I'm not looking for people who can dot 'i's and cross 't's. I need people who will hit grand slams.

Eventually, Amgen bought into the idea, and in 1993 created the Amgen Institute across two floors of renovated labs in the building that houses the OCI. As well as providing core lab running costs, Amgen pays the salaries of the six faculty members, plus two postdocs and a technician for each.

Unusually for an industry-funded facility, Mak was given almost complete freedom to explore the science that he deemed most interesting. "Tak convinced Amgen to support him in this work, more or less without direction from the company," says Nobel laureate David Baltimore, president of the California Institute of Technology in Pasadena, and one of Amgen's board of directors.

The institute's work revolves around selectively knocking out individual genes in mice to study T-cell development and other biological problems. Close observers say that Mak's uncanny ability to pick the right genetic targets for knockout experiments is the key to his success. "Tak is one of the best judges of science that I have ever seen," says Christopher Paige, vice-president of research at the OCI.

The promotion of young scientific talent is a top priority for Mak, and students at the University of Toronto flock to his lab. Graduate student Megan Cully relates what happened after she and two postdocs found a series of new links between a tumour-suppressor gene called *PTEN* and the better-known gene *p53*. "Tak immediately called

some of the top *p53* people in the world," she says. Within weeks, the experts showed up at the institute, and everyone discussed the new data.

Knockout punch

From his initial focus on T cells, Mak has branched out into studying the genes that control cell division and the 'programmed' cell death known as apoptosis. For example, his group has disrupted mouse genes including *Chk2*, which codes for an enzyme that checks for DNA damage during cell division and helps to trigger apoptosis in the event of problems². In another of Mak's hot papers, his team showed that knocking out genes that inhibit an important pathway involved in innate immunity — the initial inflammatory response to an infection — causes mouse embryos to die of liver degeneration, as a result of apoptosis³.

Mak believes the institute's success has as much to do with its approach to science as its funding from Amgen. "I urge people in the lab to do the experiments that will move the field," he says. "I'm not looking

Free thinker: Tak Wah Mak, founder of the Amgen Institute.



BABAK

for people who can dot 'i's and cross 't's. I need people who will hit grand slams."

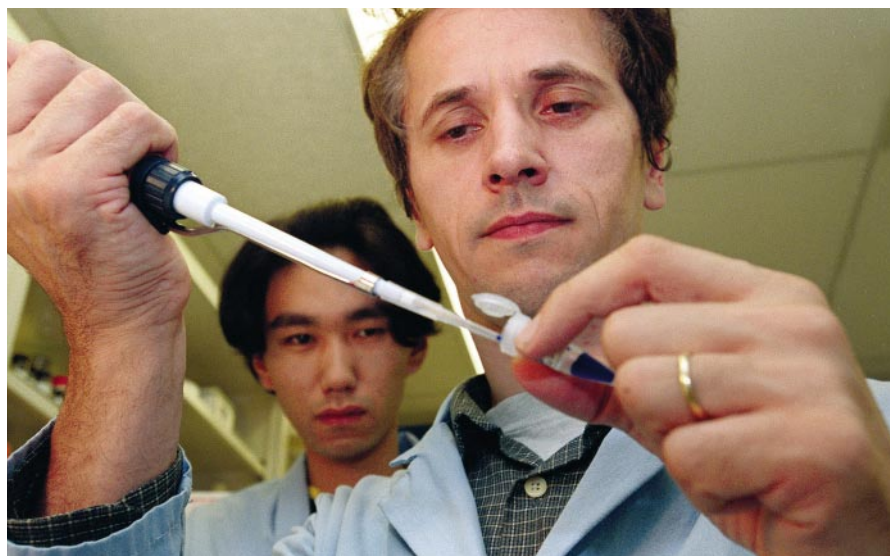
Thanks to this policy, Mak now has some serious competition as the institute's star performer. On the same floor as his lab is the workplace of Josef Penninger, an Austrian immunologist who seems to cultivate the image of an eccentric scientist. With his unruly dark hair, the gangly and engaging Penninger has become something of a celebrity, featuring on Canadian television and in magazine articles.

Three of Mak's five hot papers in 1999–2000 were co-authored by Penninger — and on two of them, the Austrian was the corresponding author. One of the joint papers showed that a protein called OPGL is involved both in the development of immune cells and in bone loss⁴. The other revealed the diverse roles of the protein TRAF-6, involved in inflammation and the maturation of cells called osteoclasts that help to repair and remodel bone⁵. In the third joint production, Mak and Penninger found that a protein involved in processing lipids regulates the activation of T cells⁶.

Penninger also scored two hot papers without his boss. One identified a novel apoptosis-inducing factor and laid out its detailed workings in a previously unknown pathway of cell death⁷. The other linked OPGL with arthritis⁸. "Josef is an exceptionally bright young guy," says Alan Bernstein, president of the Canadian Institutes for Health Research and director of the Samuel Lunenfeld Research Institute, also in Toronto. "He knows how to motivate his team and have fun."



The promotion of young scientific talent is a top priority at the institute.



Celebrated performer: Josef Penninger aims for projects that will have 'impact on people's lives'.

Penninger's success did not come overnight. He joined Mak's lab as a postdoc in 1990, and was recruited to the Amgen Institute's faculty a year after its inception. But initially, he remained in Mak's long shadow. If he continued to work on T-cell development, Penninger knew he would always be seen as a Mak acolyte. But Canadian agencies were unwilling to fund his attempts to expand into other areas. Jim Woodgett, a molecular biologist at the OCI, blames Canada's tendency towards egalitarianism: given Amgen's bankrolling of the institute, he suspects that reviewers were happier backing projects elsewhere.

Carving a niche

But the tide turned for Penninger in 1997, with the publication in *Nature*⁹ of a paper describing the knockout of a gene called *Sek-1*, which is involved in an important stress signalling pathway. After that success, funding bodies could no longer ignore Penninger's claims, and he began to investigate bone loss and heart disease, knocking out a variety of genes he thought might play important roles.

This focus on prominent diseases was grounded in practical considerations. "I am always looking for projects that will have some impact on people's lives," he says. His projects have also had a major scientific impact. Today, a line of 14 empty champagne bottles sits perched on Penninger's windowledge. Each carries a reference to one of his papers in *Nature* or *Science*.

The final member of the Amgen Institute's triumvirate of citation stars is Andrew Wakeham — who isn't even a faculty member. Creating knockout mice involves mutating a gene and then injecting it into scores of pre-implantation embryos in the hope that, in some cases, a phenomenon known as homologous recombination will cause it to displace the normal version. Wakeham is a

technician who is highly skilled at injecting embryos. Mak hired him as a research associate and he is named as a co-author on many papers emanating from the institute. "The standing joke is that Drew will beat us all," says Penninger.

But the institute can be a harshly competitive place to work. After it had been operating for three years, for instance, Mak called in an outside panel — chaired by Baltimore — to scrutinize its output. As a result, three investigators' contracts were not renewed. The move attracted some criticism, particularly after two of the outgoing researchers were replaced with postdocs from Mak's lab.

In addition to its scientific output, Mak says that the institute has provided its parent company with three drug leads — although none has yet been publicly declared. And its work on OPGL is helping to underpin Amgen's plans to develop the protein as a treatment for osteoporosis and arthritis.

But looking to the future, Mak admits to worries. Amgen's chief executive officer, Gordon Binder, and Larry Souza, the company's head of research, both retired last year. Although the institute's funding was renewed until 2008 before Souza's departure, incoming research chief Roger Perlmutter has yet to make his plans for the institute clear — and Amgen did not respond to requests to be interviewed for this article. But despite Mak's doubts about the future, other researchers would happily swap places with him. "I'd take that risk any day," says Woodgett. ■

Trisha Gura is a science writer based in Cleveland, Ohio.

1. Yanagi, Y. *et al. Nature* **308**, 145–149 (1984).
2. Hirao, A. *et al. Science* **287**, 1824–1827 (2000).
3. Rudolph, D. *et al. Genes Dev.* **14**, 854–862 (2000).
4. Kong, Y.-Y. *et al. Nature* **397**, 315–323 (1999).
5. Lomaga, M. A. *et al. Genes Dev.* **13**, 1015–1024 (1999).
6. Sasaki, T. *et al. Science* **287**, 1040–1046 (2000).
7. Susin, S. A. *et al. Nature* **397**, 441–446 (1999).
8. Kong, Y.-Y. *et al. Nature* **402**, 304–309 (1999).
9. Nishina, H. *et al. Nature* **385**, 350–353 (1997).

♦ <http://medbio.utoronto.ca/faculty/faculty.amgen.html>

e-access

These letters form part of *Nature's* current debate on access to the scientific literature. For more examples, see ► <http://www.nature.com/nature/debates/e-access/index.html>.

Free online availability substantially increases a paper's impact

Sir—The volume of scientific literature far exceeds the ability of scientists to identify and use all relevant information. The ability to locate relevant research quickly will dramatically improve communication and scientific progress. Although availability varies greatly by discipline, more than a million research articles are now freely available on the web.

Here we investigate the impact of free online availability by analysing citation rates. Online availability of an article may not greatly improve access and impact without efficient and comprehensive search services; a substantial percentage of the literature needs to be indexed by these search services before scientists consider them useful. In computer science, a substantial percentage of the literature is online and available through search engines such as Google (www.google.com), or specialized services such as ResearchIndex (www.researchindex.org)—although the greatest impact of online availability is yet to come, because comprehensive search services and more powerful search methods have become available only recently.

We analysed 119,924 conference articles in computer science and related disciplines, obtained from DBLP (dblp.uni-trier.de). In these fields, conference articles are typically formal publications and are often more prestigious than journal articles, with acceptance rates at some conferences below 10%. We estimated citation counts and online availability using ResearchIndex, excluding self-citations.

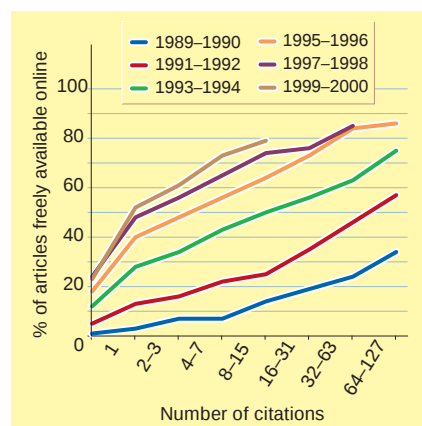
The figure shows the probability that an article is freely available online as a function of the number of citations to the article, and the year of publication of the article. The results are dramatic, showing a clear correlation between the number of times an article is cited and the probability that the article is online. More highly cited articles, and more recent articles, are significantly more likely to be online, in computer science. The mean number of citations to offline articles is 2.74, and the mean number of citations to online articles is 7.03, an increase of 157%.

We analysed differences within publication venues (the proceedings of a

conference for a particular year, for example), looking at the percentage increase in citation rates for online articles. When offline articles were more highly cited, we used the negative of the percentage increase for offline articles: hence if the average number of citations for offline articles is two, and the average for online articles is four, the percentage increase would be 100%. For the opposite situation, the percentage increase would be -100% . Averaging the percentage increase across 1,494 venues containing at least five offline and five online articles results in an average of 336% (median 158%) more citations to online computer-science articles compared with offline articles published in the same venue.

If we assume that articles published in the same venue are of similar quality, then the analysis by venue suggests that online articles are more highly cited because of their easier availability. This assumption is likely to be more valid for top-tier conferences with very high acceptance standards. Restricting our analysis to the top 20 publication venues by average citation rate gives an increase of 286% (median 284%) in the citation rate for online articles.

Free online availability facilitates access in many ways, including provision of online archives; direct connections among scientists or research groups; hassle-free links from e-mail, discussion groups and other services; indexing by web search engines; and the creation of



Analysis of 119,924 conference articles in computer science and related disciplines. The actual percentage of articles available online is greater, owing to limitations in the extraction of article information from online documents and limitations in locating articles on the web. Only points with greater than 100 articles are computed.

third-party search services. Free online availability of scientific literature offers substantial benefits to science and society. To maximize impact, minimize redundancy and speed scientific progress, authors and publishers should aim to make research easy to access.

Steve Lawrence

NEC Research Institute, 4 Independence Way,
Princeton, New Jersey 08540, USA

Authors willing to pay for instant web access

Sir—Authors of journal articles want their efforts certified by peer review and made conveniently available to the widest possible readership. They do not expect or receive royalties. What they hope for is 'impact'—attention, especially from other researchers, and recognition, especially from those who decide on hiring and promotions. Journal articles have greater impact if they are immediately and widely accessible. Maximum impact is achieved by immediate free web access (see *Nature* **410**, 1024–1025; 2001).

Journal publishers are taking advantage of the web's quick, convenient delivery of information by creating electronic versions of their traditional print journals, accessible only to subscribers or to clients of an institution with a site licence. Authors are not offered immediate free web access for their articles.

The Entomological Society of America (ESA), publisher of four leading entomological journals, recently began selling immediate free access (see <http://cssrrvr.entnem.ufl.edu/~walker/tjwbib/walker.htm>). The results suggest that a market-driven transition to free access for all articles in all journals is possible.

ESA's business plan is simple: it will provide immediate free web access, at a fair price, to authors who want it. As the cost of offering this rises (because of subscription cancellations), the price will increase. No author will be required to purchase it, and sales of subscriptions to the journals will continue as long as they are profitable. The endpoint of this plan is uncertain, but it may lead to the demise of paper publication and subscriptions, as authors and the institutions that support them embrace free access and strive to reduce costs.

Direct costs of the present system include printing paper issues, limiting access to electronic versions, and making past and present volumes accessible in hundreds of research libraries. Indirect costs are reduced impact of articles and severely restricted access by researchers in smaller institutions and in developing countries. Nonetheless, many stakeholders

believe that printed issues, or at least tolls in the form of subscriptions and site licences, will continue indefinitely.

ESA began selling immediate free web access in January 2000. During the first two months of the service, authors bought it for 13% of articles, rising steadily to 59% during March and April 2001. The price for the service is currently 75% of the price of 100 paper reprints, for example \$90 for a 7-page article. This price provides a greater profit margin than for paper reprints, which are expensive to produce and deliver. Immediate free web access requires only that the PDF file of the article is made freely accessible on ESA's web server.

If immediate free access is a profit-making service that many authors want and will pay for, why is ESA apparently the only publisher that sells it? For scientific societies, the answer is probably that their institutional inertia is great and their members have yet to lobby for it — as ESA members did. Commercial publishers may fear that selling immediate free access to those who want it may lead to all authors buying it, in which case revenues from subscriptions and site licences might cease.

On the other hand, societies have supplemented modest incomes from lower-priced library subscriptions with member dues and page charges. Without journal subscriptions, societies and commercial publishers will collect page charges to pay for refereeing, editing and composing. Publishers will pay nothing to make the articles freely web accessible, because research libraries and PubMed Central will post them without charge.

Authors should encourage publishers to provide immediate free access at a fair price. Other things being equal, many will prefer to publish in journals that provide it, especially as electronic literature indexes begin linking directly to the e-versions of articles. Most authors would like nothing better than for their articles to be available in full text, without tolls, via links in widely used literature indexes.

Thomas J. Walker

Department of Entomology and Nematology,
University of Florida, PO Box 110620, Gainesville,
Florida 32611-0620, USA

Impact factors, and why they won't go away

Sir — Brunstein in Correspondence¹ asks whether online publishing will herald the end of impact factors. Unless he is forecasting the end of print publications altogether, this is doubtful. Were print journals to disappear, however, I am confident that a new impact factor would be invented. Information scientists are

already computing web impact factors².

It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the journal impact factor as a surrogate. The latter practice is fraught with difficulties, as Seglen and others have pointed out³. As long as scientists publish articles containing lists of cited references, it will be possible to calculate impact factors. It is to be hoped that citation practices on the web will become sufficiently standardized to permit accurate calculations.

It will be necessary to distinguish between citations to URLs for research articles, on the one hand, and, on the other, to readerships as reflected in 'webometric' studies measuring web activity. One ordinarily assumes that there are many more readers than citers, but there is a widespread mythology that authors are cited more than they are read!

Eugene Garfield

Institute for Scientific Information, 3501 Market
Street, Philadelphia, Pennsylvania 19104, USA

1. Brunstein, J. *Nature* **403**, 478 (2000).
2. Bjorneborn, L. & Ingwersen, P. *Scientometrics* **50** (1), 65–82 (2001).
3. Seglen, P. O. *Br. Med. J.* **314**, 498–502 (1997).

Evolution is what's needed, not revolution

Sir — In the current debate on web availability of peer-reviewed scientific literature there seems little attempt to retain the advantages of the present system of journals.

These are, first, the element of specialization (the journal's field) and, second, the quality and novelty of the science the journal demands for publication (its place in the pecking order).

Both aspects are crucial, as authors wish their articles to be immediately read by the appropriate audience and to accrue the kudos appropriate to the significance of their work. So any universal web-based system must accommodate these two aspects.

This could be achieved if existing journals were replaced by equivalent websites that were run not by commercial publishers but by the learned societies on a non-profit basis. Provided the societies did not use their sites to generate extra income, they could keep costs to a minimum.

Reviewing could be carried out as now; the costs would be covered by page charges to authors (for both accepted and rejected submissions), and by small charges for web access to published articles. The subscription charges would be scaled: relatively high for an institution

subscribing for the benefit of all its staff, at intermediate level for an individual laboratory, and very low for a personal subscription.

Control of access could be through the predesignated IP addresses of the servers and individual terminals of subscribers. Thus an individual laboratory could have online subscription to its favourite sites for a fraction of the cost that it must now pay to get them as printed journals. If a library wanted paper on its shelves, it could either pay the site to sell it what we currently call a 'journal' or it could have a licence to print the content out as part of its institutional subscription.

A great deal of the effort for web publication (for example, generating PDF versions of text and figures in the house style of the site) can be undertaken by the authors, since they are the keenest to see the article in the public domain.

The current requirements for publication in the *Journal of Biological Chemistry*, in which every aspect is electronic, show that this is a straightforward procedure, which, while not cost-free, is not prohibitively expensive when printing costs do not have to be covered.

Colyn Crane-Robinson

Biophysics Laboratories, University of Portsmouth,
St Michael's Building, White Swan Road,
Portsmouth PO1 2DT, UK

The Net is many people's only chance of access

Sir — Pakistan is not on the publishing map. I doubt if the collective population of 140 million manage to subscribe to more than a few costly journals. On the other hand we have only about 300,000 computer users. Withholding full text from a country such as Pakistan is thus ridiculous. I would suggest that full text should be made available to everyone on the Internet. This would not affect journals financially, since most people using this service in countries such as Pakistan would never be able to subscribe to them.

F. A. Khan

1A White House Lane, Sunder Dass Road,
Lahore, Pakistan

Erratum In Stevan Harnad's Commentary on freeing the scientific literature (*Nature* 410, 1024–1025; 2001), the estimate of the minimum cost of peer review was not from the American Institute of Physics but from a summary of a group discussion by Mark Doyle of the American Physical Society. The \$500 estimate used in that discussion included only peer-review costs, not post-acceptance costs. The URL for the estimate is: <http://documents.cern.ch/archive/electronic/other/agenda/a01193/a011935t11/transparencies/>.

No occupation for a gentleman

Is the urge to humiliate an opponent the true impetus for scientific progress?

Rivals: Conflict as the Fuel of Science

by Michael White

Secker & Warburg: 2001: 417 pp. £17.99

Walter Gratzner

"There is little friendship in the world, and least of all between equals." So said Francis Bacon, often held up by historians as the first modern scientist. If you allow Michael White's thesis, then the trait of antagonism is enlarged in scientists to the point of deformity; for he believes that it is their dominant driving force. But is there really less contention in other professions, in sport, commerce or the arts? The pianist Leopold Godowsky told an anecdote about two great violinists, Mischa Elman and Jascha Heifetz. He encountered Elman during the interval of Heifetz's first sensational recital at the Carnegie Hall. Elman affected boredom with the proceedings. "Warm today," he remarked. "Not," the pianist rejoined, "for pianists." Did Heifetz gloat over his rival's discomfort, and was Elman driven to practise until his fingers ached so as to catch up? History does not reveal, but it would have been in character.

It is certainly true that White's cast of characters includes as nasty a bunch of intellectual bruisers as can be imagined. Newton, Hooke, Owen and Edison all kept their abundant reserves of bile on a hair-trigger, and pursued their perceived rivals to the grave and beyond.

White offers some fine examples of invective which we, in our age of carefully coded disparagement, can only admire. Consider this passage from Thomas Hobbes's seventeenth-century philippic, *De Corpore*: "So go your ways, you Uncivil Ecclesiastics, Inhuman Divines, Dedoctors of morality, Unasinous Colleagues, Egregious pair of Issachars, most wretched Vindices and Indices Academiaturum." Hobbes has in his sights the Oxford professors John Wallis and Seth Ward, critics of his mathematical writings. (White does not interpret, but an Issachar is one who places profit before principle, and Vindex, a vindicator or defender, was a pseudonym adopted by Ward in his attacks on Hobbes; the last epithet refers to a book by Ward devoted to a defence of the ancient universities.)

In the seventeenth century it was not uncommon for 'natural philosophers' to deposit sealed evidence of their discoveries with a learned society instead of publishing, thus safeguarding their priority without giving their rivals the opportunity of profiting from the knowledge. Sometimes the key

was concealed in an unbreakable anagram. So Robert Hooke published his law of elasticity in the form *ceiinossttuu*, which translates as *ut tensio sic vis*: stress, in other words, is proportional to strain.

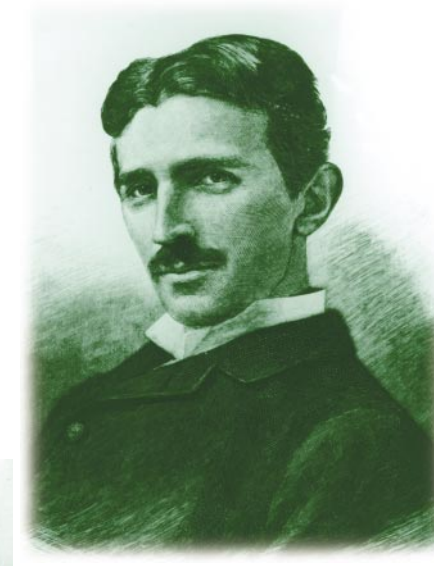
White, however, makes out that, were it not for the desire to sink an opponent, Newton would not have published his best work at all in his lifetime. This, surely, is stretching credulity. Why should Newton have been so obsessively eager to traduce Leibniz's reputation and destroy his claim to the calculus if he did not covet the full measure of glory and applause for inventing it himself? Would Darwin really have kept his theory to himself had he not been alerted to Alfred Wallace's impending publication? Darwin did not crave immortality, but he wanted recognition for his work to an extent of which he was a little ashamed. In a letter to Charles Lyell he apologized for his "trumpet feelings" and was tortured by the fear that, in publishing an abstract of his conclusions at the same time as Wallace, he might be acting dishonourably. "I would far rather burn my whole book," he concluded, "than that he or any man should think I had behaved in a paltry spirit." There have been others — Niels Bohr and Paul Dirac are well-known examples — for whom a deeper understanding was its own reward, and who took as much pleasure in the discoveries of others as in their own.

Rivalry does permeate the scientific

profession (nor is White the first to document it), but Newton and Darwin did their greatest work with no competitor in sight, and Watson and Crick set out in pursuit of the DNA structure long before they learned that Pauling was at their heels. White fails, to my mind, to sustain his proposition that competition is the engine of scientific progress, for it repeatedly emerges as a demoralizing and corrosive influence. His concluding chapter is an analysis of the motivation of the cyber-tycoons Bill Gates and Larry Ellison, but here the battlefield is not so much science (or at least technology) as commerce. And the world has drawn little advantage from the jousting of the billionaires.

White makes a clever case for the benefits that have flowed from the contest in space between the United States and the Soviet Union, but even here he is on shaky ground. Freeman Dyson, no less, has argued that the

Intellectual bruisers: Nikola Tesla (right) and Thomas Edison (below) competed for recognition.



HULTON ARCHIVE

HULTON ARCHIVE

US space programme was grievously mis-conceived from the start and that all the projects designed to assert the nation's technological virility — putting a few people into low orbits or retrieving fragments of Mars — were pointless extravagances, which did nothing to enlarge human knowledge or the limits of useful technology.

White makes some curious judgements. He accepts without reservation the (to me) unconvincing evidence that Werner Heisenberg held back the German atomic bomb project for reasons of conscience; he does not reveal Heisenberg's famous mistake in estimating the critical mass of fissile material. And he exculpates Wernher von Braun, whose activities led to the deaths of some 3,000 Londoners and many more labourers at the Peenemünde missile site; von Braun, he tells us, was "a scientist first and a weapons designer second", was no overt supporter of the Nazi party and was only doing his humble best to help his country win the war. (So that's all right, then.)

There is no dearth of mistakes, small and not so small. Max Perutz did not "unravel haemoglobin" in the year in which the DNA structure was revealed; Aaron Klug did not win his Nobel prize for genetics; Maurice Wilkins was not an X-ray crystallographer, nor did he scorn model building; Francis Crick never worked on "the effect of magnetic fields on development of cells"; the Allies did not use the Enigma machine to send misleading signals to the Germans; Martin Ryle was not the originator of the Big Bang theory and Enrico Fermi was no Hungarian physicist; Silesia is not in Czechoslovakia; and, for that matter, Eisenhower did not lose a presidential election to Kennedy.

There are, moreover, some inscrutable passages of explanation. Why are we instructed that proteins "form enzymes ... and antibodies and are integral to many huge molecules such as DNA and RNA"? And what is one to make of Maurice Wilkins's gift to Rosalind Franklin of DNA "in the *beta* [sic] or *hydrated* form", which is "colloidal, and in some respects more troublesome to work with". There is a wise maxim which states that a little inaccuracy can save a world of explanation, but here the principle has surely been carried to excess.

But I do not want to end on a disabbling note. Mistakes will no doubt be put right in the paperback edition, and *Rivals* is agreeably written, and seldom less than absorbing. You do not have to swallow the thesis on which it uneasily rests to take pleasure in these unusual, often lurid and sometimes scandalous episodes in the history of science; and you may come away with some useful tips on how to knee your competitors in the groin. ■

Walter Gratzer is at the Randall Centre for Molecular Mechanisms of Cell Function, King's College, Guy's Campus, London SE1 1UL, UK.



BRIDGEMAN ART LIBRARY

Statistical conundrum: did the tea go into the cup before or after the milk?

Still waiting for the revolution

The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century

by David Salsburg

W. H. Freeman: 2001. 340 pp. \$23.95

David Colquhoun

This is the fun side of statistics. David Salsburg's popular account of some of the great statisticians is a great read, full of anecdotes and unusual personal information. It starts around the turn of the last century with Francis Galton, Karl Pearson, W. S. Gossett ('Student' of the *t*-test) and, of course, R. A. Fisher, and continues with Egon Pearson, Jerzy Neyman and Florence Nightingale David.

Thus far, the action is all at University College London and Cambridge, but the book then spreads out worldwide. There are accounts of the life and work of many twentieth-century statisticians, from A. N. Kolmogorov's work on the axioms of probability, through I. J. Good's work on cryptography at Bletchley Park and subsequently on bayesian methods, to the many achievements of John Tukey, not least the fast Fourier transform. Along the way there are descriptions of the great (and still unresolved) wrangles between statisticians about inverse probability and the foundations of inductive inference.

The title of the book refers to a famous chapter in Fisher's *The Design of Experiments*

(1935). In this chapter Fisher illustrates the principles of experimental design by discussing the hypothetical problem of how to test the claim that it is possible to tell whether the tea is put into the cup before or after the milk. The first riveting fact I discovered from this book is that the problem is not entirely hypothetical. It is based on an actual incident at a tea party in Cambridge in the late 1920s (the author knows someone who was there).

The book is crammed with such personal anecdotes, both amusing and sometimes (thanks to Hitler and Stalin) harrowing. For example, I discovered that I. J. Good, who worked for much of his life in the United States, is the son of a Polish immigrant to London's East End who owned a well-known antique jewellery shop (Cameo Corner) near the British Museum (I bought a ring there). Salsburg does not, however, mention Florence David's predilection for cigars of churchillian proportions. I have often wondered what impression they must have made when she eventually moved to California; these days she would probably be arrested.

Salsburg's account is entirely non-mathematical, and he makes a creditable attempt at verbal descriptions of some of the great work. Nevertheless, the difficulty of conveying the ideas in words was brought home to me by the modest amount that I felt I understood in those areas I did not know about already, such as martingales (I trust his mathematical definition is more accurate than his nautical definition of this word).

When it comes to the science in the book — as opposed to history and anecdote —

Salsburg's views will not gain universal agreement. One of his recurring themes is that the old, deterministic clockwork Universe was swept aside by the "statistical revolution ... The 'things' of science are not the observables but the mathematical distribution functions that describe the probabilities associated with observations."

My problem here is that Salsburg makes little distinction between the unavoidable variability that results from the random behaviour of atoms, and avoidable variability due to errors of observation. At the level of individual molecules, atoms and subatomic particles, random behaviour is part of the physics of the system. In subatomic physics and in my own field of single-ion channels, it is quite true that what we observe are distributions. But nevertheless we can, and do, regard the (true) means of such distributions as deterministic constants. We can measure the duration of random lifetimes with high accuracy relative to the real variability of the system.

But these are specialist areas of research. In most areas, the variability is seldom of this unavoidable single-molecule sort. Most people would think it entirely reasonable that when you measure, for example, the equilibrium constant for a well-defined reaction, there is a true value. Of course there will be experimental errors in measuring this value, but the problem is essentially deterministic.

Crucial though statistical ideas are in many areas, I cannot help thinking that Salsburg exaggerates when he talks of statistics as "probably the single most important tool of biological science". What about the genome, the electron microscope and the patch clamp? If you read the great papers in my own area, for example those of Hodgkin and Huxley, of Bernard Katz, or of Neher and Sakmann, you will find very little statistics (at least of the sort that refers to measurement errors), and there was no need for any. In other areas, such as clinical trials and psychology, the use of statistics is critically important. However, one is reminded of the dangers of excessive enthusiasm for numbers when the author unblushingly tells us that "psychology developed techniques of measuring intelligence".

Most practical scientists who are dealing with large numbers of molecules that behave in an essentially deterministic way remain largely untouched by the "statistical revolution". The lady tasting tea is famous among statisticians, but none of my colleagues recognized the allusion, and very few are interested in the basis of scientific inference. This may show a lack of intellectual curiosity — or may merely reflect the amount of time they have to spend on bureaucratic activities imposed by governments and universities — but most of the time it does not hinder their science very much. It would be nice to think that the book would be read by many biolo-

gists, and indeed by the general public, but I suspect the main audience will be statisticians — amateur and professional.

Fisher's lady tasting tea was intended to illustrate the crucial importance of randomization in experimental design. This lesson has been learned well in areas such as the design of clinical trials, but is still largely ignored in laboratory sciences. Perhaps Salsburg does have a point. Studies on transgenic animals are a mainstay of research in the post-genomic era, and they can give useful results. But it is impossible to randomize gene-knockout experiments. And nobody seems to worry that this makes it impossible to do a valid significance test on the results of their experiments. Perhaps molecular biologists should read about Fisher's lady. ■

David Colquhoun is in the Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK.

Breaking down the barriers

The Ape and the Sushi Master: Cultural Reflections by a Primatologist

by Frans de Waal

Basic Books: 2001. 433 pp. \$26, £14

Christophe Boesch

Christian theology places animals on the other side of an insurmountable barrier from humans, whereas a Buddhist believes that humans are reincarnated as animals and vice versa. Christians are trained to do their utmost to preserve a large philosophical and psychological gap between animals and humans. Buddhist and similar traditions, on the other hand, attribute intentions, feelings and a mind to all living animals, including humans. As Frans de Waal puts it in *The Ape*

and the *Sushi Master*, "inspired by the pervasive human-animal dualism of the Judeo-Christian tradition, the anthropodenial, that is the *a priori* rejection of shared characteristics between human and animals, has no parallel in other religions and cultures".

Does animal culture exist? And how do scientists approach this question? The author's approach is to contrast Japanese and occidental attitudes to it. He reveals how much the philosophical background of a society affects its view of human culture. Occidental scientists believe firmly that imitation and teaching are the basis of human culture. The Japanese sushi master, however, neither teaches nor instructs his apprentice — for at least three years, the apprentice watches his master at work, without ever being allowed to practise. After this, he will prepare his first sushi, usually with remarkable dexterity. So much for teaching. We therefore need a more open definition of culture to take account of the variability of human culture. Culture, defined by de Waal, is "a way of life shared by the members of one group but not necessarily with the members of other groups of the same species ... The way individuals learn from each other is secondary, but that they learn from others is a requirement."

To illustrate his proposition that animals can have 'minds' and intention, de Waal describes Georgia, a female troublemaker in the chimpanzee colony at Yerkes Regional Primate Research Center in Atlanta, Georgia. She would take a mouthful of water as visitors were arriving, then casually mingle with the rest of the colony and wait with closed lips until the visitors came near. She would then suddenly spray them, to the accompaniment of shrieks and laughter. About this example, de Waal observes that "if Georgia the chimpanzee acts in a way that in any human would be considered deliberately deceitful, we need compelling evidence to the contrary before we say that, in fact, she was guided by different intentions, or worse, that apes have no



High culture: potato-washing by the macaque colony on Koshima Island.

intentions, and that Georgia was a mere water-spitting robot". Isn't it far more economical to assume that if two closely related species act in a similar way, the underlying mental process is also similar? If wolves and coyotes were being compared there would be immediate agreement about that. Why should we adopt another logic when comparing chimpanzees and humans?

The different philosophical approaches to animals impinge directly on de Waal's discussion of whether animals can have culture and how our cultural environment affects our scientific approach to this question. He considers it no accident that Japanese scientists were the first to record culture in animals, as "without openness to the idea of animal culture, the potato washing by monkeys on Koshima Island might never have attracted any attention". He details the fascinating observations of how Imo, a young female macaque, invented the behaviour, and how her kin and playmates learned it from her. Occidental psychologists had great difficulty in accepting this claim of cultural transmission in animals, as it threatened to remove one of the supposed barriers between humans and animals. They rejected it on the basis that the transmission process was not rapid enough to qualify as cultural transmission, and was instead individual learning in an animal.

The author is thorough and stimulating on the basic flaws in approaching the question of animal culture from a psychological angle. First, the transmission mechanism at work is irrelevant to the problem of whether culture exists in a species. Our present knowledge suggests that cultural behaviour is learnt through different and complementary mechanisms and what counts is that different cultural traits are shared within one group. Thus, there is no doubt that culture is present in animals. The interesting question is whether animal cultures rest as much on imitation as human cultures supposedly do. This question requires much investigation as, surprisingly, we know very little about how humans acquire their culture and even less about how animals do so.

Psychologists, as de Waal observes, seem to have neglected the old rule of science that absence of evidence is not the same as evidence of absence. One point put forward by psychologist opponents of animal culture was the fact that captive chimpanzees had not been seen to imitate human models. This apparent failure might be resolved, de Waal suggests, if we "level the playing field by ... testing children with ape models to see whether they would still be better imitators". An alternative experiment — presenting a human model to apes that are completely familiar with humans — has been done, and these apes turned out to be good imitators. "The surprise here is how the investigators interpreted this result. Instead of concluding that apes are a match for young children

Science in culture

Standard-bearer of culture

A new series of Italian encyclopaedias on the history of science.

Giovanni F. Bignami

In the politically black period between 1923 and 1925, industrialist Giovanni Treccani launched a new cultural era. He acquired, and donated to the Italian state, the priceless fifteenth-century illuminated bible that had belonged to Duke Borso d'Este. At the same time he founded the institution that still bears his name, the 'Istituto della enciclopedia italiana fondata da Giovanni Treccani'. Treccani's vision was to create an encyclopaedia pitched at a level somewhere between the monograph model of the glorious *Encyclopaedia Britannica* (1768) and the more abbreviated form of Larousse's *Grande dictionnaire universel* (1866).

Three-quarters of a century later, *The New York Times* described the Istituto della enciclopedia italiana as "the greatest Italian cultural institution". Its fascinating history represents one of the few examples of continuity in a country whose recent history has been marked by major discontinuities.

Work on the Treccani encyclopaedia was still in progress when, inevitably, in 1933 it became a state project, receiving the blessing of Benito Mussolini. Mussolini gave the editors total cultural freedom, the guidance of, among others, Guglielmo Marconi, and generous financial support. Understandably, he kept for himself the authorship of the entry 'fascismo'. Marconi died in 1937, just before the monumental work — 35 beautiful volumes — was presented to the king. In recognition of his achievement, Treccani received the hereditary title 'degli Alfieri' (of the standard-bearers).

Work on the *magnum opus* continued throughout the war years, with noteworthy independence. The geography section, for example, was coordinated by Roberto Almagià, a Jew who was weathering out the Holocaust years as a refugee in Vatican City.

In the 1950s, the encyclopaedia was enlarged, and the cultural activities of the Treccani institute diversified. In 1991 a special law was passed to foster "collaboration between the State and the Institute for cultural initiatives in Italy and abroad".

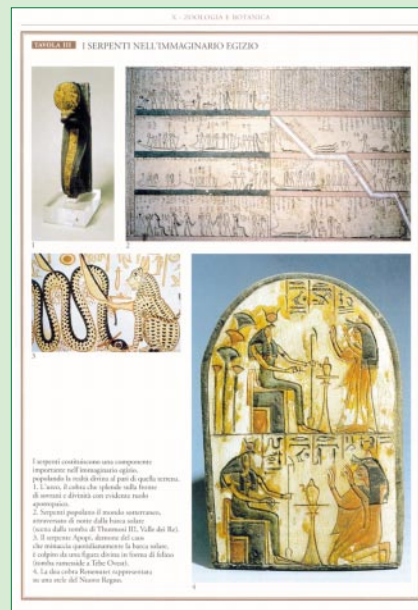
Although rigorously adhering to the Italian language (and setting high standards for it), the Treccani works have for decades been based on authoritative international contributions. The latest major work in progress is the eight-volume *Storia della scienza*, on the history of science. The first volume, on ancient science (*La scienza antica*), was published in March, and the last, on big science (*La grande scienza*), is still being written. An English-language edition is planned.

La scienza antica, with contributions from Jan Assman and Geoffrey Lloyd, is a delight. A description of an 'ophiology treatise' from Egypt tells you everything you want to know about snakes. Elsewhere, there is a translation of Archimedes' own words on the winding and unwinding of his spiral.

Treccani is the encyclopaedist's encyclopaedia. *Britannica* says of it: "One of the most important of all, the *Enciclopedia italiana* is famous for its lavish production, its superb illustrations, and its lengthy, scholarly, and well-documented articles."

Giovanni F. Bignami is in the Department of Astronomy, University of Pavia, 27100 Pavia, Italy.

► <http://www.treccani.it>



when both are equally familiar with the model, they conclude that human-reared apes are special," notes de Waal. I strongly recommend this stimulating and highly readable book, which takes us on a revelatory tour of different cultures and species. ■

Christophe Boesch is at the Max-Planck-Institute for Evolutionary Anthropology, Inselstrasse 22, 04103 Leipzig, Germany.

More primatology

Primate Origins of Human Cognition and Behavior, edited by T. Matsuzawa (Springer,

2001, DM229, \$119) includes a chapter on the 'sweet-potato-washing' monkeys of Koshima Island.

The Chimpanzees of Tai Forest by Christophe Boesch and Hedwige Boesch-Achermann (Oxford University Press, 2000; £55 (hbk), £22.50, \$39.50 (pbk))

Tree of Origin, edited by Frans de Waal (Harvard University Press, 2001; \$29.95)

Significant Others: The Ape-Human Continuum and the Quest for Human Nature by Craig B. Stanford (Basic Books, 2001; \$25).

Formally speaking

Mathematical theorems can be created by formalization of everyday expressions.

John L. Casti

It is often said that words can kill. But they can also create. Mathematics offers an excellent example of how this creative process works. Consider, for example, the common phrase, "Beauty is in the eye of the beholder". The utility of this phrase for everyday life resides in the ambiguous term 'beauty'. What does it really mean? Well, no one can say objectively, as the term is used in daily speech to mean a plethora of often contradictory things; 'beauty' is an informal notion, and therein lies its beauty (if you'll pardon the pun). But mathematicians hate the informal; it is hard to prove theorems without having a clear-cut notion of what the objects of your theory actually mean. So a lot of mathematics has been developed around just this idea: formalization of the informal.

A good illustration of this formalization process is Kurt Gödel's famous result on the incompleteness of logical systems. This result revolves about the notion of the truth of propositions about numbers. Is it 'true' that $1 + 1 = 2$? The logician Alfred Tarski showed in the 1920s that it is simply not possible to capture completely the everyday concept of 'truth' within the confines of a fixed logical system.

A large part of Gödel's genius was seeing how to replace the informal idea of truth with a concept that can be formalized, the idea of proof. Using the formal notion of proof, Gödel then showed that there must exist some statement about numbers that cannot be proved or disproved using the rules of logical deduction — but that the statement can actually be seen to be true by jumping outside the logical system and viewing the statement from a metalinguistic perspective.

There are many ways to state this result in linguistic terms. My favourite is to say

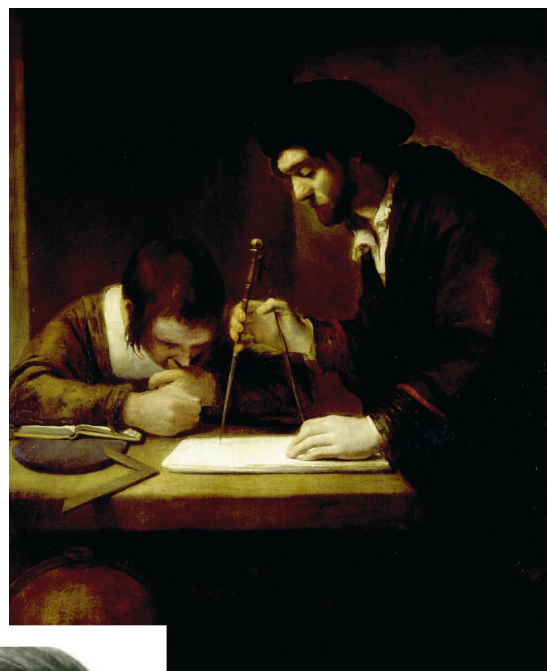
Everyday language is incredibly rich in its ability to evoke images and intuitive concepts, but to prove theorems it must be somewhat constrained.

that one must always remember that number theory is *about* numbers. In other words, Gödel's incompleteness theorem shows that there is an irreducible semantic component to mathematics. It is not just a game of shuffling meaningless strings of symbols about in accordance with a set of transformation rules. The objects of mathematics really do mean something, and that meaning cannot be completely replaced by pure syntax alone.

Here is another example of how formalization of the informal leads to new mathematics. Consider the seemingly innocent query: "What is the smallest number that cannot be expressed in words?" This statement seems to pick out a definite number. Let's call it U for 'unnameable'. But a moment's thought shows that there is something fishy about this labelling. On the one hand, we seem to have just described the number U in words. But U is supposed to be the first number that cannot be described in words! A variation of this paradox seems to have originally been suggested to Bertrand Russell by a certain Mr G. G. Berry, a librarian at Oxford University.

The Berry paradox contains its own unformalizable notion, the concept of denotation between the terms in its statement and numbers. In arriving at his results on randomness, complexity and the limitations of rule-based knowledge, Gregory Chaitin of the IBM Watson Research Center had the insight to see that the way to formalize this concept was to shift attention to the phrase: "the smallest number not computable by a program of complexity N ". This phrase can be formalized by specifying a certain computer program for searching out such a number. What Chaitin discovered was that no program of complexity N can ever produce a number having complexity greater than N . Therefore, the program of complexity N can never halt by outputting the number specified by Chaitin's phrase. This fact constitutes an algorithmic complexity version of the unsolvability of the halting problem, a computer science version of Gödel's incompleteness theorem.

In fact, the halting problem itself is the end result of a formalization process by Alan



New ideas: traditional mathematics has advanced by virtue of formalizations such as Alan Turing's (left) abstract computer.



Turing, who in the mid-1930s considered the informal idea of a "computation". Turing's way of formalizing this notion was to create an abstract computer now called a Turing machine, whose structure provides the basis for the zillions of computers in use every day around the world. If this doesn't show the power of words to create, nothing does!

The history of mathematics is filled with other examples of this sort, each formalization leading to the development of a major branch of mathematics. We can mention the notions of 'rational' decision-making and game theory, 'closeness' and topology, and 'infinity' and calculus, as illustrations of this claim. In a more contemporary vein, one can only wonder when the equally informal idea of 'complex' will be formalized into a rich mathematical theory of complex systems.

In all cases, what is involved in this process is the realization that everyday language is incredibly rich in its ability to evoke images and intuitive concepts. But to prove theorems it is necessary to constrain this expressive power somewhat, while at the same time not suffocating it entirely. Finding the right formalization is a large component of the art of doing great mathematics. ■

John Casti is at the Institute of Econometrics, Operational Res & System Theory, Technical University of Vienna, Argentinierstrasse 8, A-1040 Vienna, Austria.

BRIDGEMAN

SPL

Knife-edge of design

Jack Cohen

Victorian biologists were fond of the phrase "The profligacy of nature". As biology lost its contact with natural history and took up with Mendel and Fisher in the early 1900s, this profligacy — which had so impressed Darwin and Wallace — was replaced by Latin squares and normal distributions. The bean-counters had forgotten nature. We have to remind biology students today that two parents produce only two parents: about 14 of every 16 starlings die without breeding, a female frog lays about 10,000 eggs, of which about two breed, and all but about two of 40 million cod eggs contribute to food chains, and not to the future of their species. Many of us had forgotten this by the 1960s. John W. Saunders, in his excellent 1970 book on embryology, wrote: "Almost without fail each egg produced in the right environment does form a new individual that in turn makes sperm or eggs that begin another generation."

Reproductive biology has changed again since then. We used to believe, like Saunders, that reproduction is basically replication with few mistakes. We now realize that even bacterial replication generates many mistakes, requiring a whole suite of enzymes to repair them — a process that does not usually recover the original sequence.

I would date the beginning of the change to the discovery of p53, the 'cell-suicide' protein. Before that, examples of programmed cell death (apoptosis) in our embryology teaching were surprising exceptions. After p53 was discovered, the apoptosis bandwagon led to detailed investigation of branching apoptotic mechanisms, regulation of developmental cell numbers, mutation control,

recognition of DNA double-strand breaks, and recognition and repair of other DNA problems. Caspases and anti-caspases, the main enzymes involved, became part of cell biologists' picture of the eukaryotic cell. Rather than the cyclical, balanced machinery of the old biochemical-pathway charts, the metazoan cell came to be seen as tottering on a knife-edge between physiology and pathology. Normal physiology is not so much buffered as machine-gunned if it strays from the narrow path. Cell death is normal.

Two other revelations confronted our earlier prejudices. First, prokaryotes have 'train-wrecks' in at least a third of their replications, so a functional genome requires gene conversion and other nucleotide-replacement mechanisms. This has pushed *Escherichia coli* off its pedestal as the classical example of simple replication. We had thought that gene-conversion and chromosome-crossover mechanisms were invented by eukaryotes for the business of squeezing mendelism out of meiosis. We now see that they are rooted in the repair of ancient prokaryote (archaeal) mistakes. Second, mammalian spermatozoa have far too many double-strand breaks to contemplate human reproductive technologies such as intra-cytoplasmic sperm insertion without some concern. Woody Allen was wrong — most spermatozoa are cripples.

The 'few mutated mendelian alleles' population-genetics models of Fisher and Haldane should have been replaced by the 'recombination of very many alleles' Lewontin school in the 1960s or 1970s, when wild populations were found to carry too many genetic variants for the former. Kimura's proposal that most of the variety is 'neutral' delayed this. But measurements of wild populations of eels in eastern US rivers (elvers are much more varied than adults), for example, show that breeding organisms are usually a tiny fraction of the recombinational possibilities, regenerating the variety in their offspring. Haldane believed that loss of one-third of a population, as mutant alleles came in or were lost, would be disastrous. We now realize that 1 in 1,000 breeders is normal, and 1 in 1,000,000 is not unusual.

Sperm numbers illustrate the point well. In ascomycete fungi we can see all the products of each meiotic event: effectively, both DNA strands of the 'gamete' level are expressed in the eight ascospores. We can see that they get about a third of the products of each crossover 'wrong' in gene conversions. If metazoan crossovers do the same thing, and if these gene conversions render gametes redundant, this would exactly account for the huge numbers of sperm produced. Only a few sperm can normally reach the site of

Reproductive wastage

"Now that we have information about the genetic ecology of wild populations, and we can see how far from nature Mendel, Fisher and Haldane were, it is time to grasp the complexity of real life."

fertilization, the vast majority being labelled and phagocytosed — a conclusion that has only recently been generally accepted.

This arresting reversal of our paradigm of biological reproduction seems not yet to have been generally noticed. University population-genetics courses are still wedded to mendelian peas and laboratory *Drosophila* models, and most modern evolutionary and population-genetic models are grounded in few-mutant-allele thinking, often with a neutralist viewpoint. Now that we have information about the genetic ecology of wild populations, and we can see how far from nature Mendel, Fisher and Haldane were, it is time to grasp the complexity of real life instead of using easy arithmetic and fairy stories to describe how reproduction works.

Perhaps this proneness to mistakes in reproduction is evidence that well-designed, efficient biology is out of reach of the natural selection of undirected genetic change. We could go further and claim that terrestrial life has achieved only barely enough effective multiplication to keep biology going.

My preferred alternative is to see such variety as exuberantly exploiting any ecological possibility with a new life-form. Rather than regarding all the plaice larvae that fail to become breeders as 'waste', I wonder at the variety of successful species of flatfish, from dabs to halibut, all derived from twisted, aberrant teleosts that have adapted to lying on one side and warping their skull bones to bring the lower eye onto the top. If life had not been so prodigal, Paley might not have been so impressed by its appearance of design. ■

Jack Cohen is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

FURTHER READING

Sakkas, D. *et al.* Origin of DNA damage in ejaculated human spermatozoa. *Reviews of Reproduction* **4**, 31–37 (1999).

Ninio, J. Illusory defects and mismatches: why must DNA repair always be (slightly) error-prone? *BioEssays* **22**, 396–401 (2000).

Cox, M. M. *et al.* The importance of repairing stalled replication forks. *Nature* **404**, 37–41 (2000).

Cohen, J. Sperm galore: why mammals produce huge numbers of spermatozoa. *Science Spectra* **16**, 14–20 (1999).



Spawn free: but only a tiny fraction will survive.

Entrepreneurial insects

Stuart L. Pimm

Four British insect species have increased their geographical range as a result of climate warming. The underlying mechanisms vary from a change in food to an increase in habitat. Other species may not be quite so lucky.

Britain sits off the unfashionable fringe of a continent near the bottom of the biodiversity league¹. The size of Britain's butterfly fauna elicits expressions of sympathy from lepidopterists elsewhere. There are also occasional murmurs of encouragement — global warming might favour those species that once lived precariously, hoping for summers with enough sunshine. Will these butterflies rise to the challenge or remain provincial, hidebound by their local adaptations and so restricted at the margins of their geographical ranges to isolated subpopulations? Writing on page 577 of this issue², Thomas and colleagues show that two butterfly species, as well as two species of bush cricket, are unusually entrepreneurial at the edges of their ranges.

With so many Britons expressing so much passion for so few butterflies, Thomas *et al.* can consider butterfly ranges that have been mapped out in exquisite detail over the past two decades. The silver-spotted skipper (Fig. 1) was once found only on warm, south-facing chalk hillsides in southeast England. Now it lives even on north-facing slopes there. There is more to this result than an addition to Parmesan's catalogue³ of northward-moving butterflies. Like other butterflies, the skipper has many subpopulations that are more or less isolated from each other, reflecting the patchy distribution of suitable breeding habitats. Some patches of habitat will be too small or too isolated to support persistent butterfly populations. A warming climate has created more and larger patches. In turn, these provide stepping-stones to yet other suitable patches that were previously too remote or too small to have been occupied permanently. Because of climate warming, the total area of habitat available to the silver-spotted skipper has doubled since 1982, but the butterfly's geographical range has expanded threefold.

The distribution of the brown argus butterfly (Fig. 2) also expanded from the 1970s to the 1990s, according to two atlases that compile records of butterfly distribution across 10-kilometre-square grids. In the south of England, the range has filled in and more of these squares are now occupied than in the 1970s. In the north, the range has expanded considerably from what was previously only sporadic occurrence. As Thomas *et al.* reveal, there has also been a shift in the butterfly's choice of host plant,

from sun-loving *Helianthemum chamaecistus* to *Geranium* species. These live in habitats that were previously too cool for this butterfly. Indeed, the authors find that whereas females of stationary populations prefer to lay eggs on the same species of plant that they ate in their youth, expanding populations — irrespective of their upbringing — prefer the widely distributed geraniums. Spreading into patches of geraniums has allowed the butterfly to cross the previously impossibly large gaps in the distribution of *Helianthemum*.

Thomas *et al.* also consider two species of bush cricket that have also increased their ranges over the past 20 years. One of these, the long-winged conehead bush cricket, comes in long-winged as well as extra-long-winged forms; Roesel's bush cricket has flightless and flying forms. On the margins of these species' geographical ranges, the incidence of the more mobile forms is greatly increased. That leads, the authors assume, to a greater potential for dispersal.

Can one ever predict what will happen as climate warms? Will species march towards the warming poles or up mountainsides in an orderly fashion? Both answers are surely 'no'. Even during the post-glacial warming of the past ten millennia, species moved at very different rates⁴. The present rate of warming is much faster, so orderly changes will be even less likely. What Thomas *et al.*² have described are several synergistic mechanisms — changes in the ability of insects to move through habitat patches, changes in their food supply, and in their ability to disperse — that lead to geographical ranges expanding faster than expected. These synergies are unexpected even for species as well documented as British butterflies. For the tropics, far to Britain's south, where most butterfly species live (and live largely unknown lives), concerns about the effects of global climate change should be heightened, not allayed. The mechanisms documented by Thomas *et al.* are not only unexpected but both idiosyncratic and capricious.



Figure 2 Brown argus — shift in its choice of host plant to *Geranium* species.

Before your colleagues snatch this copy of *Nature*, ask them whether they think a warming climate will increase or decrease the density of a species' preferred habitat patches; offer more or fewer food plants; and favour more or less population dispersal. In each case, the correct answer is 'it depends'. The same answer will probably apply if one is investigating the north, middle or south of a species' geographical range. However much information is available, one might as well toss a coin as to try to predict individual range changes in detail. They depend on too many interacting factors and will be obvious only after the event.

One can make more general predictions, however. The outcomes of individual coin tosses are uncertain, but their aggregate behaviour is not. Don't bet a million pounds on the result of one toss, but go ahead and predict at least one tail from 20 tosses of a fair coin and you will be on to a winner. Metaphorically, a species with a large geographical range, typical for species



Figure 1 Silver-spotted skipper — its range has expanded threefold since 1982.

ROBERT WILSON

ROBERT WILSON



100 YEARS AGO

As a general rule, the pitch of a musical note does not in any way depend upon its intensity, but solely upon the wave-length. It appears probable, however, that any wave motion of very great intensity produces distorted effects. Thus we find that a very loud sound may so affect the ear of the observer as to appear flatter than it really is. This is a purely subjective effect... If a C tuning-fork (middle C, 256 vibrations per second) be *strongly* bowed, and then be quickly brought near the ear, before its loud note has had time to die away, the sound will appear flattened to about B_{b1}, or even A₁, the amount of the effect being different to different ears... The amount of the subjective effect differs with different individuals, both in pitch and in intensity. What to one person appears a flattening of a minor 3rd, to another auditor appears a flattening of only a major 2nd, but in every case it appears to be a flattening and not a sharpening. Also the loudness of the subjective note appears different, even to the different ears of the same person... Much of my own musical work has been done amongst male voices, and I have frequently noticed that a singer of good concert-room power may, if practising in a small room, seemingly sing with flat intonation.

From *Nature* 30 May 1901.

50 YEARS AGO

The leading article in *Nature* of April 28 on "Biologists in the Modern State" draws attention once more to the need for biologists with sound basic training in one of the physical sciences. The Agricultural Research Council is seriously concerned about the shortage of plant and animal physiologists with an adequate knowledge of chemistry and physics and of biochemists with biological experience. Posts in these subjects have remained unfilled for long periods owing to lack of suitable applicants. It is recognized by the Council that it is rarely possible in the course of study for a first degree to include adequate training in both biological and physical sciences; and hence applications for the Council's studentships from candidates wishing to study biology after graduating in a physical science, or to study a physical science more thoroughly after training in biology with a physical science as a subsidiary subject, are given very sympathetic consideration.

From *Nature* 2 June 1951.

in temperate regions, is tossing many global-warming coins. Each population might live on different plants (or perhaps suites of them), across landscapes with different structures of habitat patches, and in range centres or at range margins. And each population will make different trade-offs between its ability to disperse and other essential biological functions. In aggregate, among widespread species, there will be winners and losers, but there will always be survivors.

In the tropics, each of the many more participants in the global-warming gamble plays a far riskier game. Each outcome will be as idiosyncratic and biologically fascinating as those uncovered by Thomas and co-workers. But, because tropical species typically have smaller geographical ranges⁵,

each species bets its existence on far fewer tosses. Human actions have forced small ranges on many species, while others are born this way. The broad lesson from Thomas and colleagues' results should not be awe at how quickly a few species benefit from global change, but concern with how rapidly many may be harmed by it. ■

Stuart L. Pimm is at the Center for Environmental Research and Conservation, MC5556, Columbia University, 1200 Amsterdam Avenue, New York, New York 10027, USA.

e-mail: stuartpimm@aol.com

1. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. *Nature* **403**, 853–858 (2000).
2. Thomas, C. D. *et al.* *Nature* **411**, 577–581 (2001).
3. Parmesan, C. *et al.* *Nature* **399**, 579–583 (1999).
4. Overpeck, J. T., Webb, R. S. & Webb, T. III *Geology* **20**, 1071–1074 (1992).
5. Stevens, G. C. *Am. Nat.* **133**, 240–246 (1989).

Superconductivity

Rehearsals for prime time

Paul Grant

Superconductivity seems to have been forever waiting in the wings. Although superconducting power cables are about to go live, will the newest material, magnesium diboride, become the class act of the future?

Practical superconductors made their debut in the early 1960s. This followed the discovery of intermetallic compounds that promised high-performance electromagnets cooled by liquid helium to temperatures of about 4 K. But it took 20 years for one of the players, niobium–titanium (NbTi), to become the material of choice for superconducting wires in real applications, such as magnetic resonance imaging and high-energy physics. In 1986, a new family of high-temperature superconductors burst on the scene, based on complex copper-oxide materials. This time the potential lay in electric power applications using refrigeration technologies spanning 20 to 100 K. Despite the brittleness of the copper oxides, a viable, albeit expensive, wire technology was in place by 1992 and 'street performances' in the form of power cables are currently in production.

Then, this January, in what perhaps was the briefest billing given a modern physics discovery, *Nature* published a paper¹ reporting superconductivity at 39 K in magnesium diboride (MgB₂), an unexpected new talent for such a simple compound, which has been around since the 1950s. Now, barely five months later, some 35 physicists report in three acts^{2–4} (starting on page 558 of this issue) substantial progress towards improving the properties of MgB₂ that are vital to applications requiring high electric current and magnetic field. Superconductivity, like other players in the theatre of science, moves faster these days.

Whether or not a material is superconducting depends not only on its temperature but also on the strength of the ambient magnetic field, arising either from current flow or an applied field. In addition to having zero electrical resistance, an ordinary 'type-I' superconductor has the ability to shield itself completely from magnetic fields. But above a critical field, H_c , the superconductivity is destroyed. However, in the 1930s it was noticed that an applied magnetic field could partially penetrate some materials, eventually called 'type-II' superconductors, and the sample would remain superconducting. The magnetic flux flows through magnetic vortices, whose cores behave as simple metallic conductors, surrounded by superconducting regions. When the field is high enough the number of vortices occupies the total volume of the sample, and all superconductivity is lost. All practical superconductors are type II.

These magnetic vortices are the source of another operating limit on superconductors — the maximum superconducting current they can support. When a current flows in a type-II superconductor it produces a force on the magnetic vortices, causing them to move and creating electrical resistance through friction with the atomic lattice. But, thankfully, defects in the material can fix or 'pin' enough vortices such that they remain motionless, at least until the current and field become sufficiently strong that the vortices are 'unpinned' and so generate resistance. The behaviour of vortices therefore

defines the maximum current, J_C , and magnetic field, H^* , for a given temperature and material. It is common for J_C and H^* to start out fairly low for newly discovered superconductors, and MgB_2 was no exception^{5,6}. These quantities can be improved by introducing 'pinning defects' into type-II superconductors, but here success is based more on 'black art' than science.

The three articles published in this issue²⁻⁴ pursue several traditional routes to improving pinning used successfully in both low- and high-temperature superconductors. Eom *et al.*² explore the effects on J_C and H^* of making thin films of MgB_2 , in addition to what appears to be unintentional substitution of some of the boron by oxygen atoms. Thin films of almost all forms of superconductors, even single crystals, have higher values of J_C and H^* than their bulk counterparts. It is generally believed that the interface discontinuity between the film and substrate alone creates pinning defects. The authors present data on three film samples, all of which yielded values of J_C and H^* higher than bulk MgB_2 , the greatest increase being obtained in the 'accidentally' oxygen-doped sample. It will be interesting to see the results of more controlled attempts to affect oxygen content. In the meantime, Eom *et al.* have given us proof that the performance of MgB_2 , at least in the laboratory, can rival and perhaps eventually surpass that of existing superconducting wires.

Bugoslavsky *et al.*³, on the other hand, use proton irradiation to induce crystalline disorder in bulk samples of MgB_2 , an effect that is known to result in vortex pinning in many type-II superconductors. Such treatment can also reduce the temperature below which a material is superconducting, T_C , and there is some evidence of that in their data. But at a temperature of 20 K (roughly 50% of T_C) the reduction of J_C with applied field is much slower than in untreated samples, whereas H^* doubles on irradiation, about the same improvement observed by Eom *et al.* in their thin films. Bugoslavsky *et al.* used proton energies up to 2 mega-electron volts (MeV), as high as they had available, with penetration depths of about 50 micrometres in MgB_2 . In 1997, Krusin-Elbaum *et al.*⁷ reported huge increases in H^* in a mercury/copper oxide superconductor due to proton-induced fission of the mercury atoms. At a lower energy of 580 keV, protons colliding with ^{11}B (the most abundant natural isotope of boron) should result in fission and the emission of three 8.7 MeV alpha-particles — presumably creating more defects. Perhaps this would be an interesting 'sequel' plot for the authors to consider.

Some readers may wonder why there is all the fuss over a superconductor with a T_C of 'only' 39 K when we already have wire made from copper oxides that operate at liquid nitrogen temperatures (77 K). One good

reason is cost. High- T_C wires in use today are 70% silver by volume. Even if a silver-free technology can be developed using so-called coated conductors — thin films of superconductor deposited on metal tapes — manufacturing capital costs could be significant⁸. Moreover, a liquid cryogen, like nitrogen, is really only necessary for long lengths of alternating current (a.c.) cables, owing to large a.c. losses in both the superconductor and surrounding metallic casing. By developing lower-temperature but cost-effective MgB_2 wires, one could imagine a complete and relatively inexpensive energy-delivery system based on electricity transmitted over direct current cables (which do not have the above intrinsic losses) and cooled just by hydrogen gas at 25 K. The gas itself can be usefully consumed or stored by the end user. We could use something like this in California right now.

That is why for me the third paper, in which Jin *et al.*⁴ report a high J_C in iron-clad MgB_2 wires at 25 K and in 1 tesla fields, deserves the most applause of all. A J_C of 30,000 A cm⁻² is already almost high enough for power transmission cables. Magnesium, boron and iron are commodity elements, and the method used to produce the wires is the well-tested powder-in-a-tube technique which was developed for low-temperature

superconductors, such as NbTi, and which can be readily scaled up to high-volume manufacturing. The source of vortex pinning leading to such a high J_C is unclear at present, but is probably due to damage induced by the crushing and rolling processes used in drawing the wire (remember, I said this was a black art).

Although the dream that MgB_2 was just the first of a whole new family of materials was probably already over by the time the curtain came down at midnight at a special session of the American Physical Society meeting in March⁹, it may yet win plaudits in applied superconductivity. Still, we will have to await a few more sequels to this new work²⁻⁴ before we can predict whether and when MgB_2 will be ready to power the bright lights of Broadway and Piccadilly. ■

Paul Grant is at the Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94303, USA.

e-mail: pgrant@epri.com

1. Nagamatsu, J. *et al.* *Nature* **410**, 63–64 (2001).
2. Eom, C. B. *et al.* *Nature* **411**, 558–560 (2001).
3. Bugoslavsky, Y. *et al.* *Nature* **411**, 561–563 (2001).
4. Jin, S., Mavoori, H., Bower, C. & van Dover, R. B. *et al.* *Nature* **411**, 563–565 (2001).
5. Berlingcourt, T. G. *IEEE Trans. Mag.* **23**, 403–412 (1987).
6. Larbalestier, D. C. *et al.* *Nature* **410**, 186–189 (2001).
7. Krusin-Elbaum, L. *et al.* *Nature* **389**, 243–244 (1997).
8. Grant, P. M. *Nature* **407**, 139–141 (1995).
9. Tomlin, S. *Nature* **410**, 407 (2001).

Developmental biology

Fishing for morphogens

Stephane Vincent and Norbert Perrimon

Morphogens are long-range signalling molecules that are proposed to organize tissue patterning in animals. But their existence in vertebrates has been controversial. One suspect is now shown to fit the bill.

Throughout development, extracellular molecules tell cells where they are in the embryo, thereby guiding the formation of elaborate tissues. Fifty years ago, Alan Turing¹ proposed the 'morphogen' concept to explain how a molecule can provide spatial information. The idea is that a morphogen ('form-giving molecule') is secreted from a group of cells called an organizing centre, and then moves away. The activity of the morphogen decreases gradually as a function of its distance from the source. In this way, cells can detect where they are with respect to the organizing centre. The morphogen induces the cells to take on different fates according to their position. The existence of morphogens in vertebrates has been controversial, in part because it is difficult to prove that a given signalling molecule travels across a field of cells, instead of acting through intermediate signals. But, on page 607 of this issue², Chen and Schier confirm that Squint — a member of the transforming growth factor- β (TGF- β)

protein family — is, as suspected, a bona fide morphogen.

For some time now, members of the TGF- β family of growth factors have been prime candidates for morphogens. Indeed, studies of fruitflies (*Drosophila melanogaster*) have provided conclusive evidence that Decapentaplegic (Dpp) — a protein in the TGF- β family that controls the growth and patterning of the imaginal discs — is a morphogen. For example, in the type of imaginal disc from which wings will form, Dpp regulates the expression of two genes (*spalt* and *optomotor-blind*) differently according to the magnitude of its activity^{3,4}. Moreover, the cellular response to Dpp does not depend on interactions with other cells, and cannot propagate itself from cell to cell^{4,5}. So, Dpp acts directly on target cells, rather than by a relay mechanism through other cells, and it works in a graded manner — so fulfilling the requirements of a morphogen.

In vertebrates, a myriad of TGF- β molecules has been identified. But the evidence

that any of them act as a morphogen has been a matter of debate. Several of these molecules have been studied during the induction of mesoderm, from which tissues such as muscle develop, in the African clawed toad (*Xenopus laevis*). In this process, a signal emanates from the vegetal pole of the embryo to induce certain cells to produce mesoderm. In many respects, the TGF- β protein Activin can act like a morphogen in mesoderm induction. A high concentration of Activin induces the formation of the type of mesoderm from which the notochord will develop, whereas a low concentration induces muscle-producing mesoderm⁵. Moreover, Activin can cross a field of unresponsive cells that have been treated to block the synthesis of a possible intermediate signal⁶.

But several results have thrown into jeopardy the view that Activin behaves like a morphogen. In particular, it is unclear whether Activin really has a role in mesoderm induction. In addition, a version of Activin, tagged so its movements could be traced, was not found to diffuse from its source⁷. Meanwhile, another member of the TGF- β family, TGF- β 1, activates its receptor only on neighbouring cells, and not at a distance.

Now, Chen and Schier² provide conclusive evidence that the TGF- β family member Squint does indeed function as a morphogen in zebrafish (*Danio rerio*) embryos. Squint and a relative called Cyclops are known as Nodal signals, because they are highly related to the mouse TGF- β protein Nodal, which is involved in mesoderm induction. First, Chen and Schier injected RNA encoding Squint or Cyclops into a single cell of zebrafish embryos. They then investigated which mesodermal 'markers' — genes that are expressed specifically in particular parts of the mesoderm — were turned on at various distances from the source of Squint or Cyclops. The authors found that both proteins can, depending on their concentration, induce the expression of two sets of target genes. However, Squint and Cyclops act over different ranges, with only Squint being able to activate target genes several cells away from its source (Fig. 1). The authors then showed that no relay mechanism is required for this long-range action of Squint (Fig. 2).

So, Squint appears to act as a morphogen, but the related protein Cyclops does not. To try to understand the difference in the range of action of these molecules, the authors swapped the 'prodomains' of the two proteins. In some TGF- β molecules these regions are required as 'chaperones', to enable the proteins to fold accurately and to form intramolecular disulphide bonds. Exchanging the prodomains did not affect the ability of Squint to act over long distances, so this property must reside in the core protein. Defining the molecular basis of this ability will help us to understand how morphogens travel through tissues.

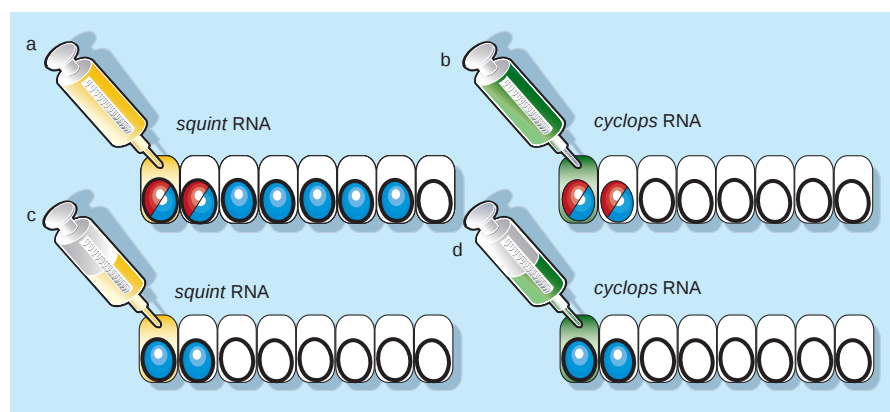


Figure 1 The Squint protein acts over long distances. To determine whether the zebrafish Squint and Cyclops proteins act as morphogens, Chen and Schier² first determined their range of action. Using zebrafish embryos consisting of 128–256 cells, the authors injected either *squint* (yellow) or *cyclops* (green) RNA into one cell and monitored the expression of target genes such as *goosecoid* (*gsc*, red) and *no tail* (*ntl*, blue). a, b, High doses of RNA injected; c, d, low doses injected. The results show that *gsc* and *ntl* are expressed in response to different levels of inducer. Moreover, at high concentrations, only Squint can induce the expression of *ntl* in distant cells. Squint therefore fulfils one requirement for a morphogen: it acts over long distances and in a graded manner.

Some insights into the mechanisms that might underlie the movement of TGF- β -like molecules have emerged from studies of *Drosophila*. Two groups^{8,9} looked at the behaviour of the Dpp protein fused with green fluorescent protein. The authors followed the distribution of the fused protein in imaginal discs, and showed that it is organized in a concentration gradient. This gradient formed quickly, was unstable, and was much steeper than the gradient of Dpp activity. Both groups concluded that the target cells help to shape the gradient. Entchev *et al.*⁹ also showed that mutations that compromise endocytosis — a process by which cells take up extracellular molecules — affect the establishment of the range of Dpp activity. These authors propose that the

long-range distribution occurs by 'planar transcytosis', in which molecules are imported into a cell, transported across it and exported on the other side.

Proteins of the Wnt and Hedgehog families have also been proposed to act as morphogens, and this has been confirmed in *Drosophila*^{10,11}. It will be important to find out how these molecules, too, might move through fields of cells. Do they travel by simple diffusion, or are they actively transported? Are extracellular cofactors involved, such as heparan-sulphate proteoglycans, which have been implicated in receiving TGF- β , Wnt and Hedgehog¹² signals at target cells? Finally, on a different note, how do cells translate quantitative information about the level of a signal into a qualitative

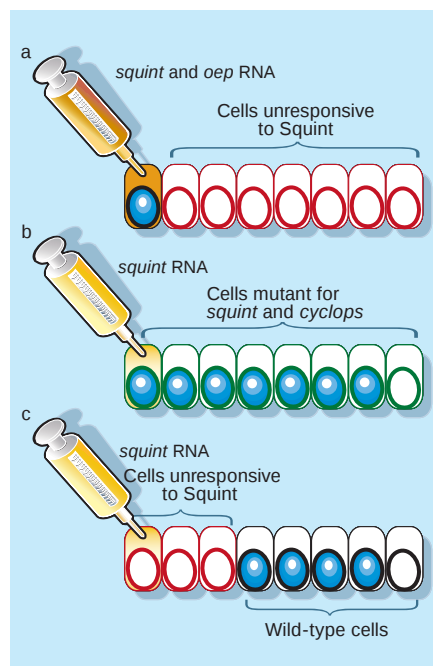


Figure 2 The Squint protein directly induces the expression of the *ntl* gene in distant cells. a, Chen and Schier² first used zebrafish embryos with mutations in the *one-eyed pinhead* (*oep*) gene. These embryos do not respond to Squint. When the authors injected both *squint* and *oep* RNA into one cell in a mutant embryo, that cell was induced to express *ntl*, but the adjacent cells (which did not express *oep*) did not. Thus, a relay mechanism that is independent of *oep* is not involved in the induction process. b, The authors injected *squint* RNA into one cell in an embryo with mutations in both *squint* and *cyclops*. Only the injected cell then expressed *ntl*. So Squint does not induce its own expression, or that of Cyclops, in neighbouring cells. c, The authors injected *squint* RNA into *oep* mutant embryos, and grafted normal cells at a distance. The intervening cells, which did not respond to Squint, did not inhibit *ntl* expression in the normal cells. So, no secondary signal is required for the long-range activity of Squint. Squint therefore fulfils the second criterion of a bona fide morphogen.

output? This fascinating problem in signal transduction must be solved before we can fully appreciate how morphogens pattern fields of cells.

Stephane Vincent is in the Department of Genetics, and Norbert Perrimon is in the Department of Genetics and Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

e-mails: perrimon@rascal.med.harvard.edu
svincent@genetics.med.harvard.edu

1. Turing, A. M. *Phil. Trans. R. Soc. Lond.* **237**, 37–72 (1952).

2. Chen, Y. & Schier, A. F. *Nature* **411**, 607–610 (2001).
3. Nellen, D., Burke, R., Struhl, G. & Basler, K. *Cell* **85**, 357–368 (1996).
4. Lecuit, T. *et al.* *Nature* **381**, 387–393 (1996).
5. Green, J. B., New, H. V. & Smith, J. C. *Cell* **71**, 731–739 (1992).
6. Gurdon, J. B., Harger, P., Mitchell, A. & Lemaire, P. *Nature* **371**, 487–492 (1994).
7. Reilly, K. M. & Melton, D. A. *Cell* **86**, 743–754 (1996).
8. Teleman, A. & Cohen, S. M. *Cell* **103**, 971–980 (2000).
9. Entchev, E. V., Schwabedissen, A. & Gonzalez-Gaitan, M. *Cell* **103**, 981–991 (2000).
10. Zecca, M., Basler, K. & Struhl, G. *Cell* **87**, 833–844 (1996).
11. Strigini, M. & Cohen, S. M. *Development* **124**, 4697–4705 (1997).
12. Perrimon, N. & Bernfield, M. *Nature* **404**, 725–728 (2000).

Earth science

Mantle cookbook calibration

Craig R. Bina

Doubts about a fundamental model of the chemistry of Earth's deep interior have now been transmuted into doubts about a standard used to calibrate these studies — gold's equation of state.

The properties of the Earth's mantle change abruptly 660 km below the surface, with sharp rises in both density and the transmission speed of seismic waves created by earthquakes. But in 1998 it was announced that highly sophisticated laboratory experiments¹ had failed to confirm this 660 km transition — instead, seeming to place the mineral transformation concerned at 600 km. Elsewhere in this issue, Chudinovskikh and Boehler², and Shim *et al.*³, describe how they have revisited the ques-

tion. The news they report is both reassuring and disturbing.

For decades, the global 660-km seismic discontinuity has been thought to coincide with the chemical reaction $\gamma \rightarrow \text{pv} + \text{mw}$, in which the mineral γ -(Mg,Fe)₂SiO₄ (also called ringwoodite or silicate spinel) breaks down to a mixture of pv-(Mg,Fe)SiO₃ (silicate perovskite) and mw-(Mg,Fe)O (magnesiowüstite, also called ferropericlae) at high pressures. This view was supported by experiments in which the reaction occurred

at pressures and temperatures believed to be appropriate to 660 km depth^{4–6}. In these experiments, samples were subjected to high pressures and temperatures inside diamond-anvil and multi-anvil devices, and were then cooled quickly and removed for analysis — a bit like looking at one's baking after removing it from the oven and allowing it to cool.

In 1998 Irifune *et al.*¹ used synchrotron X-rays to peer through the oven door and observe the baking *in situ*. But consternation greeted their report that the $\gamma \rightarrow \text{pv} + \text{mw}$ transition took place at pressures about 2 gigapascals (GPa) lower than expected, equivalent to a depth of 600 km rather than 660 km. Irifune *et al.* gave two possible explanations for this result: either their state-of-the-art pressure measurements were in error, or the common model of Earth's deep chemistry was wrong.

The two new papers^{2,3} describe experiments aimed at resolving the equilibrium pressure of the $\gamma \rightarrow \text{pv} + \text{mw}$ reaction. Using diamond-anvil devices, Chudinovskikh and Boehler² (page 574) subjected samples of Mg₂SiO₄ to high pressures and temperatures, swiftly cooled them to room temperature, and examined them using Raman spectroscopy while still at high pressures. Shim *et al.*³ (page 571) do precisely what Chudinovskikh and Boehler suggest in their closing sentence: they examined such samples *in situ*, at high pressures and temperatures, using synchrotron X-rays. The two groups used different methods of pressure calibration,

Behavioural ecology

Down on fungal farm



Leaf-cutting ants of the tribe Attini do not eat the vegetation they gather. Rather they use it to grow fungi, and then feed off the fungi. The ants' vast nests, containing

up to eight million individuals, have chambers devoted to fungus cultivation. But like all farmers, the ants face competition for the crop in the form of pathogens —

in this case other fungi, which threaten to destroy the fruits of their labours.

A couple of years ago, Cameron Currie and colleagues showed how the ants use antibiotic-producing bacteria for pest control. From observations of the species *Atta colombica* (pictured), he and Alison Stuart now describe other strategies to that end (*Proc. R. Soc. Lond. B* **268**, 1033–1039; 2001). One is simple cleanliness: the ants lick the surfaces of both the nest itself and new leaf material to keep them clear of contaminants.

The authors also sprayed fungal-cultivation chambers with spores of two pest fungi, *Trichoderma viride* and *Escovopsis*, to see the reaction. When rogue spores were found by the inhabitants, large numbers of workers congregated at the infected

site to gather up spores in their mouthparts and cart them away. If the spores reached the stage of germination, the ants turned to weeding by removing chunks of leaf along with crop and pest fungi.

The ants seem to be able to tell the difference between the two invading fungi. The specialist pathogen *Escovopsis* triggered more ants to take up cleaning duties, and their efforts went on much longer, than those directed against *Trichoderma*. But the ant farmers don't have everything their own way. Although *Trichoderma* was eliminated rapidly, *Escovopsis* hung on in the nests. It may be that this pathogen grows too quickly for the ants to keep up, or that the stickiness of the spores makes it impossible for the ants to remove 100% of them from a nest.

John Whitfield

but both conclude that the $\gamma \rightarrow \text{pv} + \text{mw}$ transition occurs at pressures and temperatures (about 24 GPa and 1,800–1,900 K) appropriate to 660 km depth. This points to problems in experimental pressure calibration, rather than to a fundamental misunderstanding of the chemistry of Earth's interior. More specifically, it seems that the source of the 2-GPa discrepancy may lie largely in the equation of state (EOS) of gold.

Our estimates of pressures and temperatures inside the Earth are indirect. They come from a variety of sources, and the consistency between them gives us confidence in their validity. But reproducing and measuring these conditions inside a device such as a diamond anvil is no easy matter. In multi-anvil devices, temperatures are measured using thermocouples, and pressures are calibrated by extrapolation from reactions at lower pressures⁶. Irifune *et al.*'s great advance¹ was to peer through the multi-anvil oven door using intense X-rays, allowing not only observation of the sample *in situ*, but also direct measurement of pressure by comparing the *in situ* specific volume of a standard material (gold) to an EOS that specifies its volume as a function of pressure and temperature. In diamond-anvil devices, temperatures are measured from radiation spectra, and pressures are determined from shifts in the fluorescence spectrum of ruby². An *in situ* approach using intense X-rays allows pressure to be calibrated from the EOS of a standard material — such as gold, or platinum³.

Problems arise in choosing an EOS for pressure calibration. As Chudinovskikh and Boehler point out, there are discrepancies between various reference EOS for gold. Indeed, experiments⁷ on the $\text{ilm} \rightarrow \text{pv}$ transition (a cousin of the $\gamma \rightarrow \text{pv} + \text{mw}$ transition but occurring in MgSiO_3 silicate ilmenite) confirmed that one such EOS⁸ (also used by Irifune *et al.* for their pressure standard) places the transition at pressures 2–3 GPa lower than does another EOS for gold⁹. The platinum pressure standard used by Shim *et al.*³, perhaps chosen to avoid the ambiguities of gold, is consistent with this second gold EOS, hence their higher pressures. In other experiments¹⁰ there has been agreement between the gold- and platinum-based pressure scales, but the group concerned appear to have used a different platinum EOS to that used by Shim *et al.*

The results of all these studies suggest that there are two distinct classes of EOS for pressure calibration — as if there were two different brands of oven thermometers (or rather barometers). One class of EOS yields pressures that place the $\gamma \rightarrow \text{pv} + \text{mw}$ transition at 660 km, whereas the other yields pressures about 2 GPa lower, placing the transition at 600 km. Furthermore, in a sort of EOS contagion, the low transition pressure for the $\gamma \rightarrow \text{pv} + \text{mw}$ reaction reported by Irifune

*et al.*¹ has already been used to calibrate pressures in a study¹¹ of the $\text{ilm} \rightarrow \text{pv}$ transition. Unsurprisingly, this yields transition pressures about 2 GPa lower than in other studies.

So the controversial 2-GPa (60-km) apparent shift of the $\gamma \rightarrow \text{pv} + \text{mw}$ transition appears to signal disagreements about high-pressure EOS of standard reference materials, rather than a misunderstanding of Earth's internal chemistry. This is both reassuring and disturbing.

It is reassuring because continued study of the transition reveals behaviour that provides a remarkably good match with the observed properties (depth, magnitude, sharpness, fine structure and topographic undulations) of the 660-km discontinuity¹². If the $\gamma \rightarrow \text{pv} + \text{mw}$ transition occurs instead at 600 km, why is no seismic discontinuity observed there, and what causes the observed 660-km discontinuity? Such paradoxes would require a radical change in mantle chemistry somewhere between 410 and 600 km deep, which is not observed geophysically.

The disturbing aspect is that, even with

the remarkable ability to look inside the oven while it is baking, there is still confusion about the relative calibrations of different pressure scales. Further interlaboratory comparisons are needed, along with assessments of the EOS from other techniques such as calorimetry, spectroscopy and *ab initio* simulations of material properties. ■

Craig R. Bina is in the Department of Geological Sciences, Northwestern University, Evanston, Illinois 60208-2150, USA.

e-mail: craig@earth.northwestern.edu

1. Irifune, T. *et al.* *Science* **279**, 1698–1700 (1998).
2. Chudinovskikh, L. & Boehler, R. *Nature* **411**, 574–577 (2001).
3. Shim, S.-H., Duffy, T. S. & Shen, G. *Nature* **411**, 571–574 (2001).
4. Liu, L. G. *Nature* **262**, 770–772 (1976).
5. Yagi, T., Bell, P. M. & Mao, H. K. *Carnegie Inst. Wash. Yb.* **78**, 614–618 (1979).
6. Ito, E. & Takahashi, E. *J. Geophys. Res.* **94**, 10637–10646 (1989).
7. Ono, S. *et al.* *Geophys. Res. Lett.* **28**, 835–838 (2001).
8. Anderson, O. L., Isaak, D. G. & Yamamoto, S. *J. Appl. Phys.* **65**, 1534–1543 (1989).
9. Jamieson, J. C., Fritz, J. N. & Manghnani, M. H. in *High-Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) 27–48 (Center Acad. Publ., Tokyo, 1982).
10. Hirose, K. *et al.* *Earth Planet. Sci. Lett.* **184**, 567–573 (2001).
11. Kuroda, K. *et al.* *Phys. Chem. Min.* **27**, 523–532 (2000).
12. Wood, B. J. *Earth Planet. Sci. Lett.* **174**, 341–354 (2000).

Human genetics

Tackling common disease

John A. Todd

Crohn's disease is characterized by inflammation and destruction of the bowel. Identification of defective variants of a gene that predispose people to the disease is an encouraging development.

The 1980s saw the identification of the genes which, when mutated, cause rare, inherited diseases such as cystic fibrosis that follow straightforward patterns of inheritance. The 1990s promised similar conquests in understanding the more common diseases (defined as more than one case per 1,000 people) that have a more complex genetic basis. The promises were not delivered.

At last, on pages 599 and 603 of this issue, two groups^{1,2} describe the first identification of a gene involved in a common disease, based on a standard approach known as linkage analysis. I predict that a steady stream of such papers will follow. The disease concerned is Crohn's disease, inheritance of which was previously correlated with³ — linked to, in genetic terminology — chromosome 16. Working independently, Hugot *et al.*¹ and Ogura *et al.*² have now pinpointed disease-predisposing variants (alleles) of the *NOD2* gene in the region of chromosome 16 that shows the strongest linkage to the disease. The bonus is that the function of the protein encoded by *NOD2* is partly known: it detects bacteria in the gut, and helps control the inflammatory response to them⁴.

Much of the genetic difference between individual people lies in differences known as single-nucleotide polymorphisms (SNPs): these are natural variations in the four bases from which DNA and genes are composed. It is largely these variations that geneticists exploit to track chromosomes from parents to children, and, in linkage analysis, to correlate the presence of certain chromosome sequences with the occurrence of disease. The segment of chromosome 16 identified as containing a gene (or genes) implicated in causing Crohn's disease was large, around 20 million base pairs. This segment contains perhaps 250 genes, variation in any one of which, or combinations thereof, could be responsible for the observed linkage.

The next task was the daunting one of finding the disease-associated gene or genes within this haystack. There are two means to this end: the 'functional candidate gene' and the 'positional' approaches, and both require some luck. In the first approach, which is based on existing knowledge of the causes of a disease, genes are selected that look likely to be involved in causation. To confirm the involvement of the candidate gene, its DNA sequence is determined in patients suffering from the condition concerned to see whether

they possess variants that might lead to an altered protein sequence or gene expression. Examples of such variants are *APOE* (in Alzheimer's disease) and *HLA-DQB1* (in type 1 diabetes).

Crohn's disease is characterized by perturbed control of inflammation in the gut and in its interaction with bacteria, and damage to the gut wall^{5,6}. In studies of how the immune system recognizes bacteria, one of the groups — Ogura *et al.* — had previously identified⁴ two intracellular versions of Toll-like receptors; these are molecules on immune cells that detect bacterial components known as lipopolysaccharides⁷. The new receptors, dubbed NOD1 and NOD2, both regulate NF- κ B, which is a master regulator of the genes involved in inflammation and has been implicated in gut-microbe interactions and inflammatory bowel disease^{5,6,8}. NOD2 was therefore a strong 'functional candidate gene' for Crohn's disease. There are, however, hundreds of plausible candidate genes in the immune system. But the real clincher for NOD2 came from its location in the genome⁴: it lies on chromosome 16, bang in the area of the strongest linkage to disease. Ogura *et al.*² had then only to sequence the gene from patients to find the disease-associated alleles.

The other group, Hugot *et al.*¹, did not know anything about the NOD2 gene when they started their quest. Rather, they began to search the 20 megabases of chromosome 16 for SNPs that were most frequently inherited with the disease. This is called association mapping. Linkage analysis is referred to as the positional approach to gene identification because it does not rely on guessing the identity of a functional candidate gene. Association mapping extends linkage analysis to map the position of the disease gene to a much higher resolution, on average to within about 60,000 base pairs. This still means that many polymorphisms must be analysed in many families, a process that is still severely limited because of the unfinished nature of the human genome sequence, the lack of complete knowledge of SNPs, and the steep cost of genotyping. Furthermore, many statistical tests were, or could have been, carried out in the search, so there is the problem of obtaining positive-looking results that are actually false, just by chance.

Hugot *et al.* did not let these limitations stop them. They took their chance, a slim one, and won. They sequenced some genomic DNA cloned from chromosome 16 that showed a sniff of disease association, identified a gene, characterized the polymorphisms in it and discovered that these were strongly associated with Crohn's disease. Their gene turned out to be NOD2. In both reports^{1,2}, the linkage result narrowed the odds.

The new work provides strong evidence

that a defective NOD2 gene makes its bearer susceptible to Crohn's disease. But it confers risk only, not certain diagnosis, and disease occurrence is dependent on the presence of bacteria in the gut and probably many other environmental factors. Moreover this is not the gene for Crohn's disease, nor even necessarily the only relevant gene in this region of chromosome 16. Other chromosomes have been linked to the disease⁹, and the genes concerned have yet to be discovered and fitted into the puzzle.

More broadly, however, these studies will encourage geneticists in their search for disease causation. It is currently assumed that common diseases might stem from many common variants with allele frequencies in the general population exceeding 15%. However, in the case of this Crohn's disease gene, NOD2, many rare variants (with frequencies of 5% or less) are involved. My guess is that the genetic basis of common disease will be a mixture of common and rare alleles, the expression of which is dependent on many environmental factors.

What about help for sufferers from Crohn's disease? Better diagnosis is a good bet. From the new findings^{1,2} it seems that up to about 15% of cases (compared with 5% in the general population) may have alleles that alter the function of NOD2. These disease-associated NOD2 alleles tend not to

be found in patients with ulcerative colitis, a related form of inflammatory bowel disease. For therapy, given that normal NOD2 protects against the gut inflammation seen in Crohn's disease, mimicking that effect might reduce the risk of disease. Sulfasalazine and glucocorticoids are among the most effective drugs in treating inflammatory bowel disease currently available and they target the NF- κ B complex¹⁰. Restoring the normal function of NOD2 in people in which the gene is defective might help prevent disease in the first place. Better diagnosis and (especially) treatments are not immediate prospects. But the new work constitutes tangible progress towards those ends. ■

John A. Todd is at the Juvenile Diabetes Research Foundation/Wellcome Trust Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 2XY, UK.

e-mail: john.todd@cimr.cam.ac.uk

1. Hugot, J.-P. *et al.* *Nature* **411**, 599–603 (2001).

2. Ogura, Y. *et al.* *Nature* **411**, 603–606 (2001).

3. Hugot, J. *et al.* *Nature* **379**, 772–773 (1996).

4. Ogura, Y. *et al.* *J. Biol. Chem.* **276**, 4812–4818 (2001).

5. Fiocchi, C. *Gastroenterology* **115**, 182–205 (1998).

6. Schreiber, S., Nikolaus, S. & Hampe, J. *Gut* **42**, 477–484 (1998).

7. Brown, P. *Nature* **410**, 1018–1020 (2001).

8. Mahida, Y. R. & Johal, S. *Clin. Exp. Immunol.* **123**, 347–349 (2001).

9. Cho, J. *Curr. Gastroenterol. Rep.* **2**, 434–439 (2000).

10. Wahl, C., Liptay, S., Adler, G. & Schmid, R. M. *J. Clin. Invest.* **101**, 1163–1174 (1998).

Chemistry

A new twist on molecular shape

Frank Weinhold

What are the forces that control the twisting and folding of molecules into complex shapes? Don't look for the answers in your organic chemistry textbook.

The forces that hinder internal rotations (twisting) about single bonds are arguably the most important factor controlling the shape and dynamics of complex polymeric molecules, including proteins, yet these forces remain widely misunderstood. Although it is relatively easy to determine the skeletal sequence of bonded atoms in a chain polymer, it is far more difficult to determine how the twists and turns at each skeletal bond conspire to produce the overall macromolecular shape. Understanding how proteins fold so quickly and correctly is widely recognized as the greatest existing challenge to structural biologists. But as the US National Institutes of Health observed last year¹: “for 50 years they've tried — and failed — to crack the code that governs folding.”

Like a three-armed turnstile, a single carbon-carbon bond is known to exert a force that opposes initial twisting from the resting state, but reverses after a 60° displace-

ment to drive the system towards an alternative resting state rotated by 120° from the first. The classic example of a simple molecule with a central carbon-carbon bond is ethane (C₂H₆). Kemp and Pitzer² first established the shape of the internal energy profile for ethane in 1936, showing that it has three energy wells (minima) corresponding to the most stable internal rotation state (conformer) shown in Fig. 1 (overleaf). The big theoretical question ever since has been: what causes the internal rotation barriers (roughly 3 kcal mol⁻¹) between one well and the next? More specifically, what causes the internal energy to rise in passing from the stable 'staggered' conformation to the unstable 'eclipsed' conformation? The emerging answer, as pointed out by Pophristic and Goodman³ on page 565 of this issue, contrasts sharply with the stock explanation presented in almost every textbook of organic chemistry.

Generations of chemistry students have

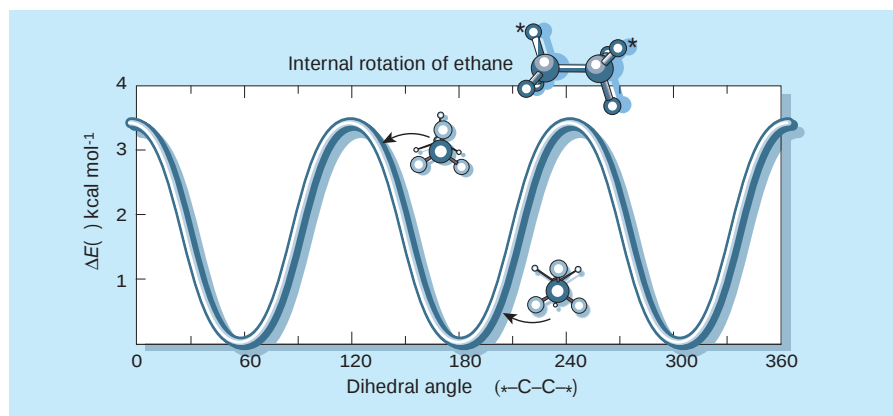


Figure 1 Energy profile for internal rotations in ethane. As one of the methyl groups (CH_3) rotates about the central carbon–carbon bond, the molecule switches three times between a high-energy ‘eclipsed’ conformation and the preferred low-energy ‘staggered’ conformation. When the molecule is viewed end-on, the near-side CH bonds appear to interleave the far-side bonds in the staggered conformations ($\phi = 60^\circ, 180^\circ$ or 300°), but cover the far-side bonds in the eclipsed conformations ($\phi = 0^\circ, 120^\circ$ or 240°). The dihedral twisting angle, ϕ , is defined as the angle between a specific pair of opposing C–H bonds (asterisks), shown here in the *gauche* conformation.

been taught that energy barriers like those found in ethane arise from spatial (steric) effects related to repulsive interactions between hard-sphere-like atoms. This explanation is intuitively satisfying, but to address the ethane problem fully, one must adopt the wave-like imagery of quantum mechanics, particularly of the orbitals that ‘house’ the electron pairs that form bonds between atoms. According to the Pauli exclusion principle of quantum mechanics, each bond orbital can be occupied by a maximum of two electrons. As pointed out in 1975 by Weisskopf⁴, two filled (doubly occupied) orbitals cannot be crowded into the same spatial region except by developing additional ripples to preserve their mutual orthogonality (a kind of ‘perpendicularity’ in the quantum space of electron orbitals). These unfavourable (energy-raising) contortions give rise to a ‘quantum pressure’ that opposes steric overlap of filled orbitals. According to the textbook picture, such steric repulsion between bonds (due to the overlap of two filled bond orbitals) is the cause of the ethane energy barrier.

Pophristic and Goodman³ use a computational method that is closely related to Weisskopf’s picture in order to evaluate quantitatively the orbital interactions in ethane. They find that the steric repulsions in ethane are only of secondary importance (in accordance with other analysis methods) and actually favour the eclipsed conformer, whereas the stable state is the staggered conformer. It is no longer valid to assume, as the textbooks do, that steric considerations play the leading role in deciding the equilibrium structure of ethane. So what controls ethane’s preference for the staggered conformation?

The true origin of ethane-like barriers can be found in another quantum-mechanical effect arising from the interactions of

electron orbitals. There is growing awareness⁵ of the importance of hyperconjugative interactions between filled and unfilled orbitals in barrier phenomena. In the case of ethane, the hyperconjugative interaction involves partial electron transfer from a nearly doubly occupied (bonding) orbital to a nearly vacant (antibonding) orbital^{6,7}. As chemistry students also learn, the bonding orbital between two atoms arises from con-

structive mixing of two atomic orbitals, and this low-energy combination is naturally chosen for occupancy by the two electrons of the bond. But the wave-mixing rules of quantum mechanics dictate that each constructive (in-phase) combination must be accompanied by the corresponding destructive (out-of-phase or antibonding) orbital.

Figure 2 shows two-dimensional and three-dimensional depictions of the orbital interactions in the staggered and eclipsed conformations of ethane. The figures show the bonding and antibonding patterns (labelled σ_{CH} and σ_{CH}^*) for two adjacent carbon–hydrogen single bonds in each conformation. The solid and dotted contours (in two dimensions) or blue and yellow lobes (three dimensions) correspond to the peaks and troughs, respectively, of the orbital waveforms. The mixing of adjacent $\sigma_{\text{CH}}-\sigma_{\text{CH}}^*$ orbitals is more favourable (for example, blue with blue, and yellow with yellow) when ethane adopts the staggered (Fig. 2a) rather than the eclipsed (Fig. 2b) conformation.

The hyperconjugative interaction in Fig. 2a leads to delocalization of negative charge, whereby the electrons in the filled orbital are able to move partially into the unoccupied orbital. It may seem counterintuitive that the electron pair in a carbon–hydrogen bond would delocalize into an adjacent high-energy antibonding orbital, the σ_{CH}^*

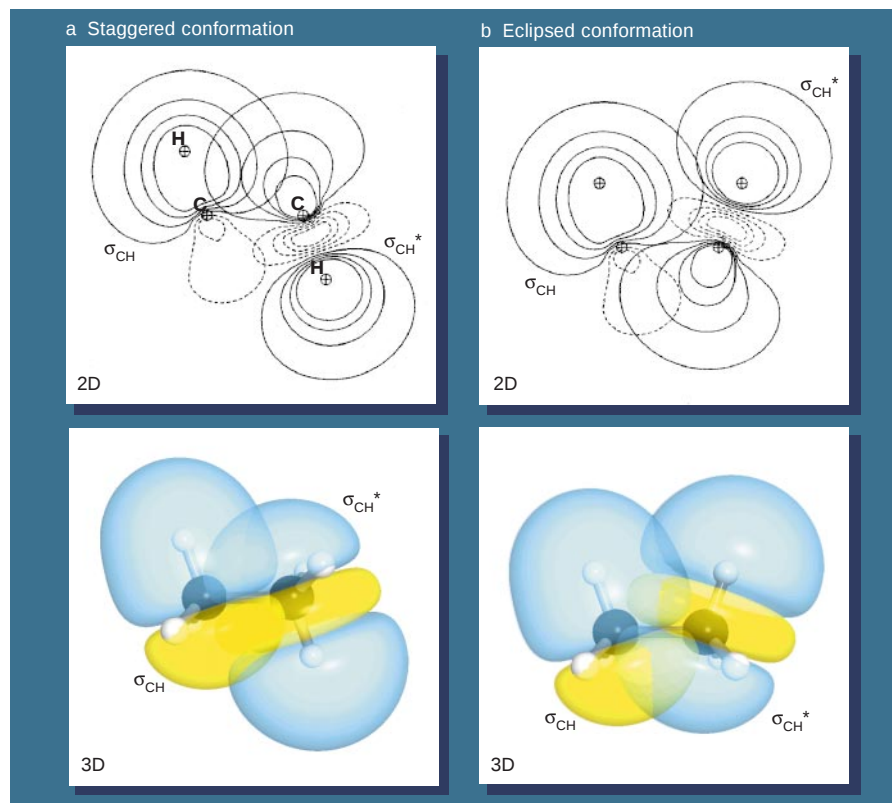
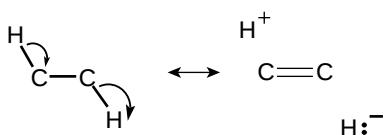


Figure 2 Contour plots (two-dimensional, 2D) and surface plots (3D) of adjacent carbon–hydrogen bond orbitals (σ_{CH} and σ_{CH}^*) of ethane. a, In the staggered conformation the more favourable orbital overlap leads to a stronger (more stabilizing) interaction. b, There is a less favourable interaction in the eclipsed conformation, leading to a higher-energy state.

orbital. Yet according to the famously weird rules of quantum superposition⁸ such hyperconjugative $\sigma_{\text{CH}}-\sigma_{\text{CH}}^*$ interactions are indeed rewarded with an additional energy lowering of a few kilocalories per mole. The enhanced strength of hyperconjugative delocalizations in the staggered arrangement is the quantum mechanical origin of the energy wells in Fig. 1.

When bonding electrons delocalize they can lead to alternative bonding patterns (resonance structures). For example, the electron delocalization associated with the $\sigma_{\text{CH}}-\sigma_{\text{CH}}^*$ interaction corresponds to a partial double-bond resonance structure, as represented in chemical notation below. Twisting about a double carbon-carbon bond is difficult, so rotation is inhibited.



So ethane's energy barriers can be viewed as a form of hyperconjugative 'resonance stabilization' for the electrons of σ -type single bonds, weaker than, but closely related to, the π -electron resonance stabilization in systems with alternating single and double bonds. From this viewpoint, ethane-like molecules adopt staggered conformations not to relieve steric congestion, but to achieve optimal resonance stabilization.

The hyperconjugative picture can also explain other conformational effects that arise in barrier phenomena. These include the 'gauche effect' (the tendency of lone electron pairs or polar bonds to adopt isomeric arrangements that are non-aligned but not *trans*), the 'anomeric effect' (*gauche* effect for carbohydrates), and various other 'stereoelectronic effects' on molecular shape and reactivity^{9,10}. Hyperconjugative interactions also account for the striking long-range effects (far beyond the range of plausible steric influences) that are observed in the barrier potentials of toluene and related aromatic species¹¹. As shown by Pophristic and Goodman, hyperconjugative interactions are also largely responsible for inducing the skeletal distortions, such as carbon-carbon bond elongation, that accompany internal rotations. At the orbital level, a host of apparently unconnected effects appear to be due to the same electronic interaction. Electrons, after all, know only a few tricks.

The hyperconjugative interactions in Fig. 2 offer new electronic opportunities for controlling the conformational destiny of complex biomolecules. Broader awareness of these interactions among molecular scientists should inspire new chemical and biochemical strategies to control twisting processes — for example, through cooperative coupling to other delocalizing inter-

actions. Improved understanding of the electronic origins of molecular twisting may also spur overdue efforts to incorporate more realistic quantum-mechanical effects into molecular modelling and protein folding calculations. The most pressing question raised by Pophristic and Goodman is: when will chemistry textbooks begin to serve as aids, rather than barriers, to this enriched quantum-mechanical perspective on how molecular turnstiles work? ■

Frank Weinhold is in the Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA.

Immunology

Brief encounter

Raymond M. Welsh

To a certain extent, the response of immune cells to infection is preset. T cells seem to be activated early on or not at all, and, once stimulated, embark on a complete programme of division and specialization.

When a doctor palpates a patient's abdomen or the glands under the jaw, the intent is often to assess the size of the spleen or the lymph nodes, which swell in response to infection. This swelling results from the proliferation of antigen-specific T and B cells, which also become specialized in different ways to control intruding pathogens. When the infection is quelled, the immune response is silenced, and many of the T and B cells die. Others lurk as slowly dividing 'memory' populations in lymphoid and non-lymphoid organs, until activated by the same or a related pathogen. Puzzlingly, the timing of the T-cell response is often more or less the same regardless of the pathogen, and T cells may continue multiplying for days after the antigen is cleared. In contrast, the magnitude of the response can vary markedly, depending on the antigen and its form of delivery (a point not lost on vaccinologists). Writing in *Nature Immunology*, van Stipdonk and colleagues¹ and Kaech and Ahmed² resolve the puzzle: they show that T cells are programmed to multiply and differentiate after only a brief encounter with a stimulating antigen.

In a typical immune response, a pathogen is first processed by antigen-presenting cells (APCs), which display on their surface short peptides from the pathogen, in complex with 'major histocompatibility' (MHC) proteins. Naive T cells may recognize this peptide-MHC complex, and this interaction starts a series of events inside the T cells, which is amplified by the interaction of 'co-stimulatory molecules' on the T cells and the APCs³. Thereafter, the T cells both secrete and respond to growth and differentiation fac-

e-mail: weinhold@chem.wisc.edu

1. *The Structures of Life* NIH Publication No. 01-2778 (National Institutes of Health, Washington DC, 2000). <http://www.nigms.nih.gov>
2. Kemp, J. D. & Pitzer, K. S. *J. Chem. Phys.* **4**, 749 (1936).
3. Pophristic, V. & Goodman, L. *Nature* **411**, 565–568 (2001).
4. Weisskopf, V. F. *Science* **187**, 605–612 (1975).
5. Michl, J. & West, R. *Acc. Chem. Res.* **33**, 821–824 (2000).
6. Brunck, T. K. & Weinhold, F. *J. Am. Chem. Soc.* **101**, 1701–1709 (1979).
7. Reed, A. E. & Weinhold, F. *Isr. J. Chem.* **31**, 277–285 (1991).
8. Weinhold, F. *J. Chem. Ed.* **76**, 1141 (1999).
9. Pettito, P. & Lerner, L. in *The Anomeric Effect and Associated Stereoelectronic Effects*: ACS Symposium Series (ed. Thatcher, G. R. J.) 539 (Am. Chem. Soc., New York, 1993).
10. Bretscher, L. E. et al. *J. Am. Chem. Soc.* **123**, 777–778 (2001).
11. Lu, K.-T. et al. *J. Chem. Phys.* **102**, 6787–6805 (1995).

tors, and take on functions that enable them to ward off the pathogens, for example by killing infected cells.

Van Stipdonk *et al.*¹ questioned how long T cells must be exposed to antigen for this response to be stimulated and maintained. They found that CD8⁺ T cells (so called because they express a molecule called CD8 on their surface) required as little as 2 hours of exposure to APCs to initiate several rounds of division and differentiation into cell-killing (cytotoxic) T cells. To show this, the authors engineered mouse fibroblast cells to express a co-stimulatory molecule, as well as a peptide from the ovalbumin protein that could stimulate a population of naive, ovalbumin-specific CD8⁺ T cells. They then treated the T cells with a fluorescent dye, the intensity of which decreases by half with each cell division, and exposed them to the fibroblasts (the APCs) for different time periods. They studied the T cells to see how much they had proliferated, how much dye they had lost (an indicator of how many times they had divided), and whether they had differentiated.

Cell division did not begin until at least 24 hours after stimulation, but by 48 hours the cells were proliferating rapidly, dividing every 5 to 6 hours. This remarkably fast generation time is consistent with previous studies^{4–6}. Moreover, when the authors separated the T cells from the APCs after as little as 2 hours, the entire proliferation and differentiation programme continued¹. The authors obtained similar results *in vivo*.

One implication is that T cells 'decide' very early during an infection whether to respond to the pathogen. This sheds light on a recent report by Mercado *et al.*⁴, who showed that the magnitude of the response

of CD8⁺ cytotoxic T cells in mice to a peptide from an intracellular bacterial pathogen, *Listeria monocytogenes*, was similar whether or not the mice had been treated with an antibiotic 24 hours after infection. The antibiotic markedly reduced the number of bacteria, yet the response remained the same. Mercado *et al.* also transferred dye-labelled T cells, specific to *L. monocytogenes*, into mice; similar proportions of the T cells divided whether the infection was allowed to continue or was blocked by an antibiotic after 24 hours. Their results again indicate that once a T cell is stimulated it usually undergoes a complete programme of proliferation and specialization.

Kaech and Ahmed² add further insight into how the magnitude of a T-cell response relates to antigen load. They infected mice with different doses of *L. monocytogenes* that had been engineered to express a peptide from lymphocytic choriomeningitis virus (LCMV). They then injected the mice with dye-labelled T cells specific to the LCMV peptide. The T cells either followed the whole proliferative programme or did not divide at all; the higher the antigen dose, the more T cells responded. If the T cells were stimulated with antigen just briefly *in vitro*, they continued to divide and differentiate² (T cells may need to divide just once to be able to differentiate in this system⁷). All of these data^{1,2,4}, coupled with the observation that the peak in a T-cell response is often sharp rather than prolonged over several days⁴⁻⁶, suggest that if a naive T cell is not stimulated during the first few days of infection, it may not respond at all.

Kaech and Ahmed also found that some of the stimulated T cells later became memory cells. The number of CD8⁺ memory T cells has been shown to be proportional to the magnitude of the T-cell response before it is silenced⁸. These results are relevant to vaccination, as the development of a pronounced memory response could depend on the stimulation of a large proportion of naive, antigen-specific T cells early during infection.

T cells are clearly preprogrammed to some extent. It seems that, once stimulated briefly, T cells undergo at least 7 or 8 divisions, with generation times between 5 and 8 hours (refs 1, 2, 4), and that 12 or more divisions can take place before the immune response is silenced⁵. However, it is unlikely that the cells are programmed so precisely that they are oblivious to the effects of other external factors. Kaech and Ahmed² found that the division of CD8⁺ T cells was blocked when interleukin-2, the receptor for the T-cell growth factor, was inhibited; it is noteworthy that CD4⁺ T cells, the functions of which are compromised in AIDS, are rich sources of interleukin-2. Moreover, the persistence of antigen can continue to stimulate T-cell proliferation up to a point, but too

much antigen can drive T cells into 'clonal exhaustion'⁹.

Are any other functions of T cells programmed events? The mechanisms behind the death of T cells at the end of an immune response, and the slow division of memory T cells, have defied understanding. But these processes might similarly be explained by the existence of some autonomous, preprogrammed clock within the T cells themselves. Raymond M. Welsh is in the Department of Pathology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA.

e-mail: raymond.welsh@umassmed.edu

1. van Stipdonk, M. J. B., Lemmens, E. E. & Schoenberger, S. P. *Nature Immunol.* **2**, 423–429 (2001).
2. Kaech, S. M. & Ahmed, R. *Nature Immunol.* **2**, 415–422 (2001).
3. Chambers, C. A. & Allison, J. P. *Curr. Opin. Cell Biol.* **11**, 203–210 (1999).
4. Mercado, R. *et al.* *J. Immunol.* **165**, 6833–6839 (2000).
5. Selin, L. K., Vergilis, K., Welsh, R. M. & Nahill, S. R. *J. Exp. Med.* **183**, 2489–2499 (1996).
6. Murali-Krishna, K. *et al.* *Immunity* **8**, 177–187 (1998).
7. Oehen, S. & Brduscha-Riem, K. *J. Immunol.* **161**, 5338–5346 (1998).
8. Hou, S., Hyland, L., Ryan, K. W., Portner, A. & Doherty, P. C. *Nature* **369**, 652–654 (1994).
9. Moskophidis, D., Lechner, F., Pircher, H. & Zinkernagel, R. M. *Nature* **362**, 758–761 (1993).

Cell biology

Channels as enzymes

Michael D. Cahalan

The TRP family of ion channels is proving rather strange. The latest quirks of behaviour include a new enzymatic activity and second messenger, the ability to conduct magnesium ions, and an involvement in cell survival.

Simply put, ion channels are proteins that form pores in cellular membranes, conducting ions into or out of cells. Once thought to be involved merely in sending electrical signals from place to place in the body, ion channels are now known to have many functions in all types of cell. They allow a variety of positive or negative ions to enter or exit cells, and are regulated by many different molecules. The TRP channels, encoded by about 20 distinct genes, are possibly the strangest ion channels. For example, some are actually enzymes, catalysing not just the movement of ions, but also chemical reactions that couple to signalling and metabolic pathways within cells. On pages 590 and 595 of this issue, Scharenberg and colleagues^{1,2}

reveal yet more unusual features of TRP channels.

The TRP-channel family is undergoing a nomenclature upheaval at present, but all parties agree that three subfamilies can be distinguished on the basis of sequence gazing^{3,4}. Each receptor folds up to crisscross the plasma membrane six times, with the amino and carboxy tails inside the cell. The transmembrane portions contribute to a cation channel, and the two ends have other functions. It is likely that four subunits assemble to form a functional channel, and that diversity is increased because different combinations of subunits can form the channels.

Members of the 'long TRP channel' (LTRPC) subfamily have unusually long amino and carboxy termini. The carboxy

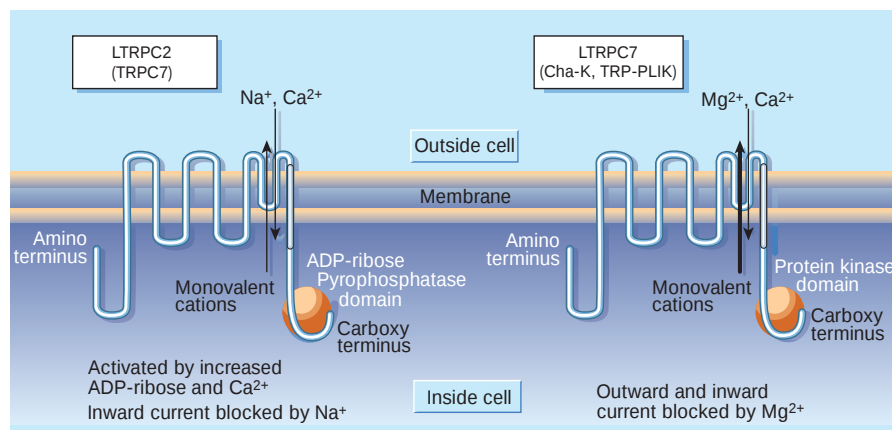


Figure 1 The behaviour of two ion channels from the long-TRP-channel family. The alternative names of LTRPC7 and LTRPC2 (refs 1, 2) are shown, along with their enzymatic activities and ion-permeation properties. The inward current through LTRPC7 is carried by calcium or magnesium ions, and is small compared with the outward current carried by monovalent cations — in other words, LTRPC7 is a strong outward rectifier¹. LTRPC2 allows inward current (carried by the combination of sodium and calcium ions) or outward current (carried by monovalent cations) to flow equally².

Daedalus

Four feet, one mouth

In days gone by, by-products were a valuable asset. The LeBlanc process for manufacturing sodium carbonate survived for decades on its by-products. These days they merely complicate marketing; demand for main products and by-products can fluctuate unpredictably. But one industry, farming, is stuck with by-products. Every liver brings one heart, unavoidably. And now that foot-and-mouth disease has made even farmers question their livelihood, Daedalus hopes to exploit modern biochemistry.

He points out that stem-cell research allows any number of cells of any sort to be made. Cells very close to their origin can develop into anything — new brain cells, for example. And by sticking with non-human tissue, all sorts of ethical dilemmas can be avoided. Daedalus recalls how rapidly a damaged liver can regenerate. Suppose, he says, a liver culture were fed with stem cells influenced by signalling chemicals and environment to develop as liver cells. You could haul an endless liver slowly from the dish, overcoming the problem of having a by-product.

So DREADCO biochemists are exploring Daedalus' theory that stem cells can be made in bulk by developing new cells in a very 'young' medium. In due course it should be possible to haul a cow leg, say, or a multi-jointed leg, from a dish whose cells are primed to develop in a leggy manner. Much of the by-product problem would be overcome, although there would still be a fixed amount of skin (for shoes) and bone (for soup and gelatine film). Ultimately all animal husbandry, a cruel and inefficient process, will be replaced by painless tissue culture.

Plants will be exploited too. Daedalus reckons that rhubarb grows because its stalk gains glucose diffusing down from its leaf. If so, an endless rhubarb stalk could be made by pulling a natural one while feeding it glucose. Again, the by-product problem (poisonous leaves) would be neatly overcome. The DREADCO team is now exploring the animal and vegetable kingdoms in search of suitable species. Rhubarb is possible; so is potato (a large tubular tuber could be pulled from the photosynthesizing fronds), and so are snake and eel, although most animals are badly designed for the job. Ultimately biomass, now seen as an answer to our energy needs, could take its place at the dinner-table as well. But probably synthesized liver, leg joints and rhubarb pie would be more welcome than pure masses of animal or vegetable cells. David Jones

termini contain domains that have enzymatic activity (Fig. 1). LTRPC7, for example, has an unusual protein kinase domain, which can catalyse the transfer of phosphate from adenosine triphosphate (ATP) to several substrates⁵, including LTRPC7 itself.

Although the natural activators of LTRPC7 remain unknown, Runnels *et al.*⁵ conclude that its kinase activity is strictly required for its channel function. They found that the addition of ATP to the inside of cells increased the flow of ions through the channel, suggesting that ATP (and, by inference, the kinase domain) is needed to open the channel.

This conclusion is challenged in one of the new papers¹. The discrepancy centres on the relationship between Mg^{2+} ions and ATP. Nadler *et al.*¹ find that the channel's activity is suppressed by Mg^{2+} ions, whether free or bound to ATP. They suggest that the addition of ATP by itself would lower the concentration of free Mg^{2+} ions. So, the channel activity observed by Runnels *et al.* in the presence of ATP might be explained by a fall in the concentration of free Mg^{2+} , without needing to invoke a role for LTRPC7's kinase activity. Other results, obtained with a hydrolysis-resistant analogue of ATP², provide further evidence against phosphorylation-dependent gating of LTRPC7. But this is a debate that is likely to rumble on.

As the authors show in the second paper², the long TRP channel LTRPC2 also has an enzymatic domain at its carboxy tail. This domain enables the protein to remove the terminal ribose-5-phosphate moiety of ADP-ribose. ADP-ribose also activates the channel, suggesting that, in this case, the enzymatic activity might actually be involved in opening or closing the channel. Once activated, the channel can remain open for a prolonged period even after ADP-ribose has been removed — a molecular memory worth pursuing further.

The LTRPC7 channel needs to be regulated properly for cells to survive: when Nadler *et al.*¹ either overexpressed or knocked out LTRPC7 in different cell types, the cells died. Under normal conditions, LTRPC7 is a strong outward rectifier¹, meaning that the inward current (carried by Ca^{2+} or Mg^{2+} ions) is smaller than the outward current (carried by monovalent cations, such as K^+). It seems plausible that, when LTRPC7 is overexpressed or knocked out, the levels of Ca^{2+} , Mg^{2+} or possibly other polyvalent cations might become unbalanced. Moreover, both LTRPC2 and LTRPC7 might be affected by ischaemic conditions, which result from oxygen depletion. Intracellular ATP-bound Mg^{2+} blocks both inward and outward currents through LTRPC7 (ref. 1), and Na^+ selectively blocks inward currents through LTRPC2 (ref. 2). In response to ischaemia, levels of Na^+ inside cells would rise and levels of Mg^{2+} — ATP would fall. So ischaemia might block

inward currents through LTRPC2, and activate inward and outward currents through LTRPC7.

These papers^{1,2} might also have a bearing on one of the major puzzles in ion-channel biology: the identity of 'store-operated' Ca^{2+} channels. When a cell's Ca^{2+} stores are emptied in response to some signal, Ca^{2+} enters the cell, because Ca^{2+} channels in the plasma membrane open. The Ca^{2+} -release-activated Ca^{2+} (CRAC) current found in the immune system is one of the best-characterized store-operated currents, yet the molecular identity and mechanism of activation of the CRAC channel have been elusive. However, TRP4 and CaT1, which are both members of the TRP family, are leading contenders^{6–8}.

LTRPC2 and LTRPC7 are also expressed in immune cells^{1,2}, and Nadler *et al.* and Perraud *et al.* describe currents that resemble those through these channels in some types of immune cell. These results imply that the conditions used to investigate CRAC and other store-operated channels in immune cells might have unintended effects on cellular physiology, by perturbing cellular ions and metabolites during whole-cell recording. My group recently characterized single channels in human T cells activated under conditions that deplete Ca^{2+} stores passively (using strong Ca^{2+} buffers). These channels resemble CaT1 channels in some aspects and LTRPC7 channels in others⁹. The molecular identity of CRAC channels is yet to be pinned down, but further characterization of single channels and the development of molecules that block specific channels would help in the investigation.

As a final note, TRP channels are important in disease as well as health. For example, the levels of specific LTRPC-family members change during the progression of tumours such as melanoma and prostate cancer^{10,11}. Before we can understand the role of these channels in disease, we will need to have a clearer understanding of how they function as molecular devices. Here, too, the development of selective channel blockers — both as experimental tools and as potential treatments — should be a high priority. ■

Michael D. Cahalan is in the Department of Physiology and Biophysics, University of California, Irvine, California 92697-4561, USA.
e-mail: mcahalana@uci.edu

- Nadler, M. J. S. *et al.* *Nature* **411**, 590–595 (2001).
- Perraud, A.-L. *et al.* *Nature* **411**, 595–599 (2001).
- Harteneck, C., Plant T. D. & Schultz, G. *Trends Neurosci.* **23**, 159–166 (2000).
- Clapham, D. E., Runnels, L. W. & Strübing, C. *Nature Rev. Neurosci.* **2**, 387–396 (2001).
- Runnels, L. W., Yue, X. & Clapham, D. E. *Science* **291**, 1043–1047 (2001).
- Philipp, S. *et al.* *J. Biol. Chem.* **275**, 23965–23972 (2000).
- Freichel, M. *et al.* *Nature Cell Biol.* **3**, 121–127 (2001).
- Yue, L., Peng, J.-B., Hediger, M. A. & Clapham, D. E. *Nature* **410**, 705–709 (2001).
- Fomina, A., Fanger, C. M., Kozak, J. A. & Cahalan, M. D. *J. Cell Biol.* **150**, 1435–1444 (2000).
- Duncan, L. M. *et al.* *J. Clin. Oncol.* **19**, 568–576 (2001).
- Wissenbach, U. *et al.* *J. Biol. Chem.* (in the press).

Obituary

G. N. Ramachandran (1922–2001)

There are few instances where research from a comparatively obscure laboratory, far removed from the main centres of international research, has made a lasting impression on world science. The work of G. N. Ramachandran, the renowned crystallographer and molecular biophysicist (or structural biologist in the current parlance), is one such rare example. Ramachandran died at Chennai, formerly Madras, on 7 April 2001. It is also remarkable that he produced so much seminal work even after withdrawing from mainstream research while still quite young.

Ramachandran had his early education at Ernakulam, now in the southern Indian state of Kerala, and earned his first degree in physics at St Joseph's College, Trichy, in Tamil Nadu state. He joined the Indian Institute of Science, Bangalore, in 1942 to study electrical engineering. But he soon came under the spell of the Nobel laureate, C. V. Raman, and — to the consternation of the professor of electrical engineering — shifted to the physics department. After earning his MSc and DSc under the supervision of Raman, he moved to the Cavendish Laboratory, Cambridge, UK, on a scholarship, where he took his second doctorate. He returned to the Indian Institute of Science in 1949, working as an assistant professor in the physics department until 1952.

It was his move to start a physics department at Madras University, at the comparatively young age of 29 for a full professor, that marked the beginning of an extraordinarily productive phase in Ramachandran's career. This period saw his greatest achievements, including determination of the structure of the protein collagen, and was characterized by what I can only think was a unique relationship between a senior university official and a young scientist. A. L. Lashmanaswamy Mudaliar (here on the left, with Ramachandran in the centre), the university's vice-chancellor, treated Ramachandran almost as a son and provided him with full administrative and financial support. Ramachandran's mandate was to develop a world-class research centre at Madras, which he duly did.

Twenty years on, however, he found it difficult to adjust to the new regime when Mudaliar retired. In 1971, he returned to the Bangalore Institute to establish the molecular biophysics unit, which today remains a major research centre. He largely withdrew from mainstream scientific research in the mid-1970s, and



Renowned crystallographer and structural biologist

subsequently worked on mathematical philosophy and logic until he was incapacitated by bad health.

It was discussions with J. D. Bernal, who visited Madras in 1952–53, that led Ramachandran to the problem of the structure of collagen. Collagen is the fibrous protein found in skin, bone and tendon, and it was to long defy attempts to solve its structure using X-ray fibre diffraction and modelling. By the early 1950s Linus Pauling had already predicted the existence of the famous α -helix — a characteristic structural feature of proteins — using simple modelling, and Watson and Crick's structure of DNA was shortly to appear. More than one group were at that time working on collagen, but Bernal felt that the problem was still open.

Following his advice, Ramachandran and Gopinath Kartha set to work, and within a couple of years developed the three-stranded, coiled-coil, helical model of collagen. This was followed by what in retrospect appears to be needlessly acrimonious controversy over the details. Ramachandran and his colleagues continued to refine the structure by including water molecules. Some 50 years on, the Ramachandran model has stood the test of time.

The controversy over the structure had two effects on Ramachandran, one negative and one positive. On one hand, he felt that there was undue hesitation in according him the credit due to him. On the other, it led him to an intense study of the core issue in the debate, the minimum possible distance between two non-

bonded atoms. This investigation eventually resulted in the celebrated Ramachandran map, proposed in 1963 by Ramachandran, C. Ramakrishnan and V. Sasisekharan. The map sets the limits imposed on polypeptide chain conformation by the need for non-bonded atoms to keep out of each other's way.

Eventually the Ramachandran map turned out to be much more broadly applicable than had perhaps been originally intended. Today it provides the simplest complete description of protein conformation. It is also the most important tool for the validation of protein structure and, in a way, has immortalized Ramachandran. But he and his colleagues' research on proteins other than collagen, and on nucleic acids and polysaccharides, was also monumental. And Ramachandran was among the pioneers who laid the foundations of the thriving field of molecular modelling.

His early work in optics, X-ray diffraction and diffuse scattering stood Ramachandran in good stead when dealing with problems in X-ray crystallography. He illuminated the field with imaginative theoretical ideas as much as he suggested practical solutions. His 1956 contribution on the use of anomalous dispersion in phase determination in X-ray structure analysis is a landmark in the development of the field. He initiated studies on crystallographic statistics. And his work on Fourier methods, along with others such as R. Srinivasan, created part of the foundations of modern X-ray crystallography. Finally, he contributed to developments in three-dimensional image reconstruction, an aspect of his work that is insufficiently appreciated. It is fitting that the last major award he received, in 1999, was the Ewald Prize of the International Union of Crystallography. This prize is awarded only every three years, and went to Ramachandran for his huge overall influence on crystallography.

To scientists of my generation in India, Ramachandran was a source of scientific and personal inspiration. Most of his contributions were based on simple, yet striking, ideas; and he showed how international science could be influenced even from less well-endowed neighbourhoods. Ramachandran left structural biology and mainstream research about a quarter of a century ago. Yet his presence in the area remains as vibrant as ever.

M. Vijayan

M. Vijayan is at the Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560 012, India.

e-mail: mv@mbu.iisc.ernet.in

Natural selection and resistance to HIV

A genotype that lowers susceptibility to HIV extends survival at a time of peak fertility.

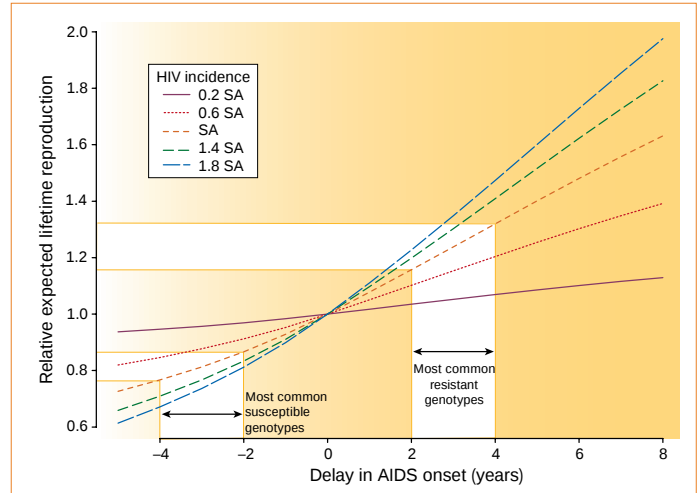
Although infectious disease is assumed to be an important cause of natural selection in humans, strong selection in favour of alleles that confer resistance to disease has been demonstrated only in the case of malaria¹. Here we show that the high prevalence of human immunodeficiency virus (HIV) in some African countries creates conditions for strong natural selection on genetic variants that affect the time between HIV infection and the onset of AIDS. Moreover, a rapid evolution of resistance to AIDS onset is expected in these populations if the current HIV epidemic conditions persist.

In 1999, the prevalence of HIV infection in adults was 20% in South Africa, 36% in Botswana, 25% in Zimbabwe and 20% in Zambia². Assuming that the current incidence of HIV infection and the mortality caused by AIDS do not change, the lifetime risk of dying of AIDS for a boy who was 15 years old in 1999 is over 65% in South Africa and almost 90% in Botswana².

Variant forms of CCR5, a cell-surface receptor for mediator molecules known as chemokines, have been found to affect susceptibility to HIV infection and the period before the onset of AIDS in people carrying alleles for the variant forms. These genetic polymorphisms will be subjected to considerable natural selection because they increase survival time during the period of peak fertility. The HIV-resistant haplotypes that are most common in Africans³ delay post-infection progression to AIDS by 2–4 years and, conversely, common HIV-susceptible haplotypes accelerate AIDS onset by a similar amount.

Figure 1 shows the estimated expected lifetime reproduction relative to the normal genotype as a function of delay in AIDS onset. The 'SA' curve corresponds to the estimated HIV incidence in 1997 in South Africa⁴. The expected lifetime reproduction is roughly 15% greater for genotypes in which AIDS onset is delayed by 2 years, and roughly 30% greater for genotypes that

Figure 1 Expected lifetime reproduction of carriers of CCR5 mutations relative to the normal CCR5 genotype as a function of the period between HIV infection and the onset of AIDS. Curve 'SA' was calculated using HIV-incidence values estimated for a rural district in South Africa⁴; other curves were calculated with all incidence values multiplied by the indicated factor. Calculations are based on 1980 demographic data⁵ (assumed to be before the AIDS epidemic) for South Africa and an exponential survivorship curve for the time until AIDS onset. Mean time to AIDS onset is 7 years for the normal genotype and changes by the times shown for resistant or susceptible genotypes. Further details will be published elsewhere.



delay onset by 4 years. The expected lifetime reproduction of HIV-susceptible (AIDS-accelerating) genotypes is reduced to a comparable extent.

Estimates of the incidence of HIV infection have a high degree of uncertainty, so we have included other curves in Fig. 1, representing different overall incidence levels. The '0.6 SA' curve, for example, corresponds roughly to incidence values estimated for a Ugandan population with a similar prevalence distribution to that for the South African population⁵. The expected lifetime reproduction for this incidence distribution varies from 9% greater than the normal genotype for an onset delay of 2 years, to 19% greater for an onset delay of 4 years.

Table 1 shows the projected evolution of haplotype frequencies and the effect on average time to AIDS onset when these haplotypes are exposed to HIV at South African incidence levels⁴. In 100 years, the frequency of the AIDS-delaying haplotype increases from 0.4 to 0.53 and the average time to onset increases by 1 year.

The selective advantage calculated here is comparable to that of malaria resistance

in heterozygous carriers of the haemoglobin S allele (which is responsible for sickle-cell anaemia in SS homozygotes) in regions where malaria is common. Our estimate of the intensity of selection caused by HIV today is also comparable in magnitude to the 30% advantage conferred by the $\Delta 32$ mutation in CCR5 (a 32-base-pair deletion that leads to loss of the receptor on lymphoid cells) in European populations over the past 700 years⁶. Our results offer some insight into the cause of this selection. Bubonic plague has been proposed to be the selective agent for the $\Delta 32$ mutation⁶. In its worst outbreak, the plague caused 25–35% mortality over much of Europe during the Black Death from 1346 to 1352. The mortality in Africa caused by HIV will rise to at least that level and probably higher.

If heterozygous carriers of this deletion had only partial resistance to plague, selection would have been of roughly the same intensity as in populations with a high prevalence of HIV infection. However, unpublished calculations (by P.S.) show that epidemics of plague persisted for too short a time for a selective advantage of 10–30% to account for the current high frequency of the disease. Instead, heterozygous carriers of the CCR5 mutation would have to have been completely resistant to plague (or to a similar pathogen), which would confer a much greater selective advantage.

Paul Schliekelman, Chad Garner, Montgomery Slatkin

Department of Integrative Biology, University of California, Berkeley, California 94720, USA

e-mail: pdschlie@socrates.berkeley.edu

1. Allison, A. C. *Br. Med. J.* **1**, 290–294 (1954).

2. Joint United Nations Programme on HIV/AIDS (United

Table 1 Projected changes under HIV-induced selection

Timescale (years)	Resistant-haplotype frequency	Susceptible-haplotype frequency	Average time to AIDS onset (years)
0	0.40	0.20	7.8
20	0.43	0.17	8.1
40	0.46	0.15	8.3
60	0.49	0.13	8.5
80	0.51	0.12	8.6
100	0.53	0.10	8.8

Estimates for African populations were calculated using standard equations for evolution of age-structured populations⁷. Predicted frequencies of haplotypes subjected to constant HIV incidence are shown. Incidence values are those calculated in ref. 4. The initial frequencies correspond to those found earlier³, with the eight most common types being collapsed into three types. The estimated frequency of the 'normal' haplotype remains almost constant at 0.4 and is not shown. The average time between HIV infection and AIDS onset is also calculated.

- Nations, Geneva, 2000).
- Gonzalez, E. et al. *Proc. Natl Acad. Sci. USA* **96**, 12004–12009 (1999).
 - Wilkinson, D., Abdool Karim, S. S., Williams, B. & Gouws, E. *J. Acqui. Imm. Def. Syndr.* **23**, 405–409 (2000).
 - Gregson, S., Donnelly, C. A., Parker, C. G. & Anderson, R. M.

- AIDS* **10**, 1689–1697 (1996).
- Stephens, J. C. et al. *Am. J. Hum. Genet.* **62**, 1507–1515 (1998).
- Charlesworth, B. *Evolution in Age-Structured Populations* (Cambridge Univ. Press, Cambridge, 1980).
- Census (International Programs Center, Population Division, United States Census Bureau, 2000).

Climate change

Increasing shrub abundance in the Arctic

The warming of the Alaskan Arctic during the past 150 years¹ has accelerated over the last three decades² and is expected to increase vegetation productivity in tundra if shrubs become more abundant^{3,4}; indeed, this transition may already be under way according to local plot studies⁵ and remote sensing⁶. Here we present evidence for a widespread increase in shrub abundance over more than 320 km² of Arctic landscape during the past 50 years, based on a comparison of historic and modern aerial photographs. This expansion will alter the partitioning of energy in summer⁷ and the trapping and distribution of snow in winter⁸, as well as increasing the amount of

carbon stored in a region that is believed to be a net source of carbon dioxide⁹.

During oil exploration of the United States Naval Petroleum Reserve no. 4 in northern Alaska in 1948–50, low-altitude oblique photographs of exceptional clarity were taken at thousands of locations between the Brooks Range and the Arctic coast¹⁰. In July of 1999 and 2000, we took photographs at 66 of the same locations spanning an area 400 km (east to west) by 150 km. We analysed pairs of new and old photographs for changes in the three principal deciduous shrubs, dwarf birch (*Betula nana*), willow (*Salix* sp.) and green alder (*Alnus crispa*), and for changes in treeline white spruce (*Picea glauca*) along the southern edge of the study area.

In 36 of the 66 repeat photo-pairs, we found distinctive and, in some cases, dramatic increases in the height and diameter

of individual shrubs, in-filling of areas that had only had a scattering of shrubs in 1948–50, and expansion of shrubs into previously shrub-free areas (Figs 1, 2). At tree-line sites, there was a marked increase in the extent and density of the spruce forest (Fig. 2). In some cases, shrub-dominated vegetation that covered about 10% of the landscape in 1948–50 had doubled by 2000. In the 30 photo-pairs in which the amount of deciduous shrubs had not increased, there was no detectable reduction in shrub abundance either.

The increase in shrub abundance appears to have been mainly the result of the growth and expansion of alder, perhaps partly because dark-coloured alder are the most conspicuous shrubs (Fig. 1). However, in several photographs ($n=4$), birch and willow were also seen to have increased in abundance. All three species belong to the same functional group and respond to experimental warming and fertilization in a positive manner⁵. This indicates that the abundance of the smaller dispersed birch and willow found throughout tussock tundra may also be increasing, and so our detection of change could be conservative. These smaller shrubs comprise most of the shrub biomass in the study area.

Our study area is in a location where human and natural disturbances (leading to successional changes) are minimal, so we attribute much of the increase in the abundance of shrubs to the recent change in climate. During the Early Holocene, warming in the Alaskan Arctic was accompanied by one or more large-scale shrub invasions¹¹, and today shrub abundance increases along latitudinal temperature gradients¹². These findings, combined with our observations, show that the vegetation of the region is able to respond to changes in climate, perhaps rapidly. The extensive peat deposits¹³ are evidence that the region has been an important sink for global carbon in the geological past. The increased primary production inferred from our photographic analysis could be a significant contributor to changes in the high-latitude carbon budget, as well as contributing to important changes in the exchange of surface energy.

Matthew Sturm*, **Charles Racine†**,
Kenneth Tape‡

*US Army Cold Regions Research & Engineering Laboratory–Alaska, PO Box 35170, Fort Wainwright, Alaska 99703, USA

e-mail: msturm@crrel.usace.army.mil

†US Army Cold Regions Research & Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755, USA

‡Geophysical Institute, University of Alaska–Fairbanks, PO Box 757320, Fairbanks, Alaska 99775, USA

- Overpeck, J. et al. *Science* **278**, 1251–1256 (1997).
- Serreze, M. C. et al. *Climatic Change* **46**, 159–207 (2000).
- Epstein, H. E., Walker, M. D., Chapin, F. S. & Starfield, A. M. *Ecol. Applic.* **10**, 824–841 (2000).

Figure 1 The Ayiyak River (N68° 53', W152° 31'), showing an increase in the density of shrub patches, the growth of individual shrubs and an expansion of shrubs into areas that were previously shrub-free. A and B denote the same locations in the old and new photographs.

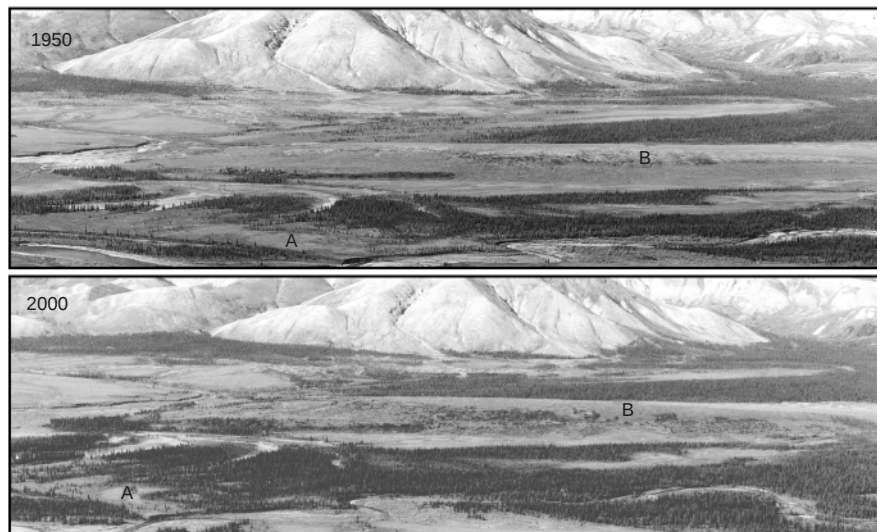
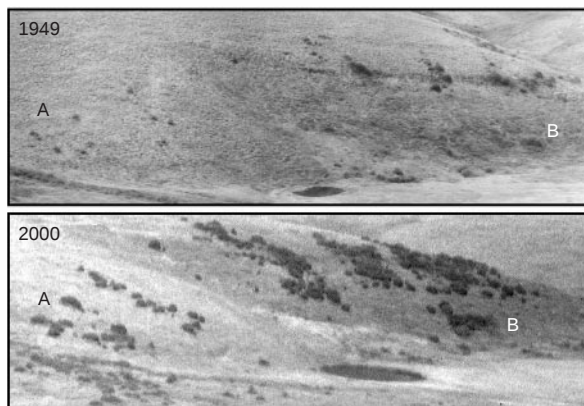


Figure 2 The Kuguruk River (N68° 06', W161° 31'), showing in-filling of spruce stands (A) and increased abundance of shrubs in the middle ground (B); A and B denote the same locations in the old and new photographs.

4. McGuire, A. D. *et al.* *Glob. Change Biol.* **6**, 141–150 (2000).
5. Chapin, F. S. *et al.* *Ecology* **73**, 694–711 (1995).
6. Myrneni, R. B. *et al.* *Nature* **386**, 698–702 (1997).
7. Chapin, F. S. *et al.* *J. Climate* **13**, 2002–2010 (2000).
8. Sturm, M. *et al.* *J. Climate* **14**, 336–344 (2001).
9. Oechel, W. C. *et al.* *Nature* **406**, 978–981 (2000).
10. Reed, J. C. *US Geol. Surv. Prof. Pap.* **302** (1958).
11. Anderson, P. M. & Brubaker, L. B. in *Global Climates Since the Last Glacial Maximum* (ed. Wright, H. E. Jr) 386–400 (Univ. Minnesota Press, Minneapolis, 1993).
12. Bliss, L. C. & Matveyeva, N. V. in *Circumpolar Arctic Vegetation* (eds Chapin, F. S. *et al.*) 59–89 (Academic, San Diego, 1992).
13. Billings, W. D. *Quat. Sci. Rev.* **6**, 155–177 (1987).

Sensory adaptation

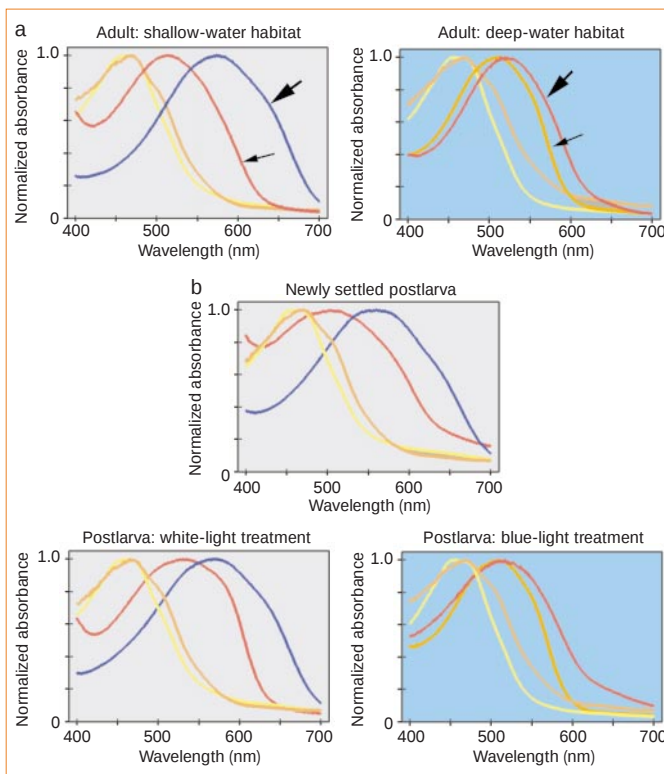
Tunable colour vision in a mantis shrimp

Systems of colour vision are normally identical in all members of a species, but a single design may not be adequate for species living in a diverse range of light environments. Here we show that in the mantis shrimp *Haptosquilla trispinosa*, which occupies a range of depths in the ocean, long-wavelength colour receptors are individually tuned to the local light environment. The spectral sensitivity of specific classes of photoreceptor is adjusted by filters that vary between individuals.

Colour vision in mantis shrimps involves up to 16 types of visual pigment^{1–3}. Spectral sensitivity is further tuned in an unusual way: transparent, coloured filters are placed in front of four classes of receptors^{3–5}. Most mantis shrimp species occupy a narrow depth range, and the spectral properties of the filters vary with their characteristic depth. Species in shallow water with bright, broad-spectrum illumination have filters that tune some photoreceptors to very long wavelengths, with sensitivity peaking at wavelengths greater than 600 nm. As sea water selectively attenuates long-wavelength light⁶, these receptors would be ineffective at depths exceeding 10 m. In deeper-living species, corresponding filters transmit shorter wavelengths, providing a better match to the prevailing light conditions^{3,7}. Animals that live in either environment may thereby perceive colour throughout almost the entire spectral range of available light.

H. trispinosa, however, lives at depths ranging from barely subtidal to 30 m or more, and long-wavelength photoreceptor classes that are useful near the surface would not function in deep water. To meet this challenge, different individuals within the species could express different sets of visual pigments, or filters could vary between individuals. At Lizard Island, Australia, we collected adults of *H. trispinosa* at depths of 1 m and 15 m and characterized the visual pigments and filters in their retinas using microspectrophotometry^{1,5}. Deep and shallow populations had

Figure 1 Normalized, average absorption spectra ($n = 2–9$) of all filter classes in *H. trispinosa* retinas. The colour of each line represents how the filter class appears to the human eye. The spectral maxima of visual pigments in the receptors that underlie each filter class are (described using filter colours in the upper-left panel): yellow, 508 nm; orange, 537 nm; red, 539 nm; blue, 558 nm. **a**, Filters in adults from shallow water (1 m) and deep water (15 m). The two longer-wavelength classes, which appear red (thin arrows) and blue (thick arrows) in retinas of shallow-water individuals, absorb shorter wavelengths in deeper-living adults. **b**, Filters of newly settled postlarvae and of juveniles after 3 months in white or blue light.



identical visual pigments (data not shown), but the filters that tune the long-wavelength receptor types (Fig. 1a, arrows) were significantly shifted towards shorter wavelengths in deeper-living animals. Thus, their underlying receptors could discriminate hues within the bluer spectrum present in deep water. We calculate that, despite a 96% reduction in illumination at wavelengths longer than 575 nm, the tuning of long-wavelength receptors of animals living at 15 m allows them to capture incident photons at 75% of the rate observed for surface dwellers.

We investigated whether the occurrence of different filter sets can be influenced by conditions during development. Newly metamorphosed postlarvae of *H. trispinosa* live in shallow water and have filters like those of shallow-water adults (Fig. 1b). We reared postlarvae under laboratory lighting, either in blue light lacking wavelengths longer than 550 nm or under broad-spectrum white fluorescent light. After 3 months, individuals reared in white light retained the filter classes of shallow-living adults, whereas the filters of the blue-light group were characteristic of deep-water populations (Fig. 1b).

To our knowledge, this is the first demonstration in any species of a tunable colour-vision system that responds directly to the light environment. The response involves precise changes in a unique system of spectral filters. Although some fish, amphibians and crustaceans vary their spectral sensitivity by changing their visual-pigment chromophores, this is controlled

by environmental temperature or photoperiod^{8,9}. Sensitivity also changes ontogenetically in some fish¹⁰, perhaps by hormonal control. In mantis shrimps, tuning could be regulated by the overall level of illumination, which decreases with depth, or could be controlled by the spectral quality of incident light.

Colour vision in *H. trispinosa* is influenced to a surprising extent by developmental conditions. Individuals living at different depths have different colour perception and may rely on different biological signals to compensate for this. In any case, these mantis shrimps express an impressive degree of phenotypic plasticity in tailoring their visual systems to their habitats. We suspect that intraspecific variation in visual function may be widespread among species that occupy variable light environments.

Thomas W. Cronin*, Roy L. Caldwell†, Justin Marshall‡

**Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA*

e-mail: cronin@umbc.edu

†Department of Integrative Biology, University of California, Berkeley, California 94720, USA

‡Vision, Touch and Hearing Research Centre, University of Queensland, Brisbane 4072, Australia

1. Cronin, T. W. & Marshall, N. J. *Nature* **339**, 137–140 (1989).

2. Marshall, J. & Oberwinkler, J. *Nature* **401**, 873–874 (1999).

3. Cronin, T. W., Marshall, N. J. & Caldwell, R. L. *Phil. Trans. R. Soc. Lond. B* **355**, 1263–1267 (2000).

4. Marshall, N. J. *Nature* **333**, 557–560 (1988).

5. Cronin, T. W., Marshall, N. J., Caldwell, R. L. & Shashar, N. *Vision Res.* **34**, 2639–2656 (1994).

6. McFarland, W. N. & Munz, F. W. *Vision Res.* **15**, 1063–1070 (1975).

7. Cronin, T. W., Marshall, N. J. & Caldwell, R. L. *Vision Res.* **34**,

- 279–291 (1994).
 8. Tsai, A. T. C. & Beatty, D. D. *Science* **195**, 1358–1359 (1977).
 9. Suzuki, T., Arikawa, K. & Eguchi, E. *Zool. Sci.* **2**, 455–461 (1985).
 10. Hawryshyn, C. W., Arnold, M. G., Chaisson, D. J. & Martin, P. C. *Visual Neurosci.* **2**, 247–254 (1989).

Developmental biology

Lungfish dental pattern conserved for 360 Myr

Lungfish, the closest living relatives of four-limbed animals, are unique in that adults lack marginal teeth and have to rely on palatal dental plates for crushing food. We have discovered that an identical pattern of tooth development is used to shape these plates in the hatchlings of fossil and living lungfish species that are separated by 360 million years (Myr) of evolution, even though the adults have very different dental forms; the same pattern is also evident in the transient marginal dentition, despite being functional only until the juvenile stage. This remarkable finding indicates that developmental programming for dentition in lungfish is uniform, unique and conserved for all tooth fields.

Adult lungfish have extensive, continuously growing tooth plates on the palate and the inner side of the lower jaws, which are formed without shedding any teeth. This type of dentition is developed and maintained through continual addition of new teeth labially and dentine from within. In this way individual teeth, arranged in radial rows, are consolidated into dental plates without loss through shedding¹. This is in contrast to the conventional marginal linear rows of teeth that form the osteichthyan dentition (including tetrapods), in which tooth shedding usually occurs through lingual development of new teeth in each position, with loss of the old tooth. The specialized dentition unique to lungfish is one

example of a strongly conserved structure that is retained through constraints in developmental patterning.

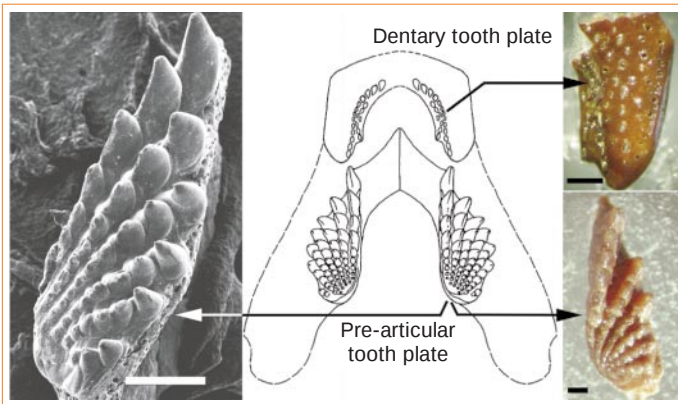
This type of development is shown clearly in the lower jaws of hatchlings of the extant *Neoceratodus* and the Late Devonian *Andreivichthys*, two lungfish taxa that are separated by 360 Myr. Superbly preserved growth series of *Andreivichthys* have been uncovered in central Russia². These include not only thousands of specimens, but also the earliest hatchling stages, the only such examples in the current fossil record.

This discovery allowed us to compare the Devonian form with similar developmental stages of *Neoceratodus*. Although the dental plates in adult *Neoceratodus* are non-toothed surfaces for crushing, they develop from toothed rows³, with new teeth being added labially to each row of the tooth plate as in adult *Andreivichthys*³ (Fig. 1, left arrow).

In addition, hatchlings of both taxa develop a marginal dentition, in which the feeding mode is different⁴ from that in the adults. This marginal dentition is also organized in both taxa as small tooth plates and is quite unlike that in other osteichthyans. Each marginal set is formed from several rows of joined teeth in the Devonian form (Fig. 1, top right) and from paired single rows in the extant lungfish (Fig. 2). This provides substantial evidence that in lungfish the mode of dentition patterning is consistent in each tooth field, whether marginal or palatal, and that it is unique and specialized. This suggests that it is evolutionarily distinct from that of all other osteichthyans.

We also observed a pattern of selected dental loss through resorption of the marginal dentition from hatchlings to juveniles. In *Neoceratodus* the sequence of loss starts medially, first with the single symphyseal tooth, after which those on the dentary are resorbed, one tooth at a time, until the mar-

Figure 1 Dorsal views of dentary and pre-articular tooth plates in the Late Devonian *Andreivichthys epitomus*. The diagram shows a reconstructed outline of the lower jaw in occlusal view and the relative positions of the two types of plate. As in the extant lungfish (Fig. 2), the dentary tooth plate (top right; PIN 1302; arrow originates from and points to the centre of the radiating rows)



is completely resorbed during early ontogeny, and the dentary bone is lost, whereas the pre-articular tooth plate (left and bottom right; PIN 1328, 1548) continues to add new teeth. Both comprise radial rows of teeth. Left, scanning electron micrograph of the pre-articular plate, taken at an angle to show the new teeth at the labial end of each row (arrowhead). Six tooth rows can be seen, all showing a gradation from small, worn medial teeth to larger new teeth; rows 1–4 have 10 teeth, row 5 has 6 teeth and row 6 has 5 teeth. Material courtesy of N. Krupina. Scale bars, 500 μ m.

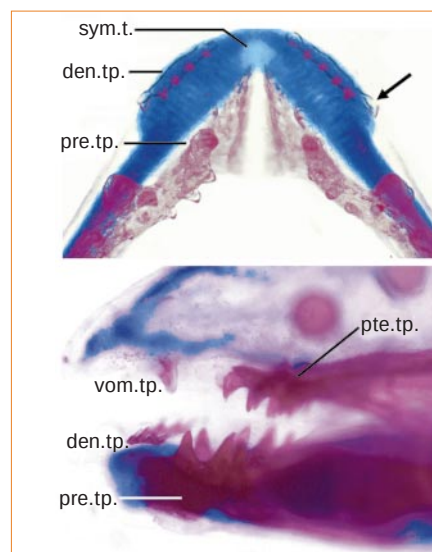


Figure 2 Tooth plates of the extant *Neoceratodus forsteri*. Images show whole mounts of a hatchling at stage 56, stained red for bone and dentine, and blue for cartilage. Top, occlusal view of lower jaw in which resorption of dentary teeth has begun to occur (one medial tooth from left side). The symphyseal tooth has been completely lost, as shown by the pale blue region. Arrow shows addition of new teeth, which are not yet attached, at the postero-lateral ends of the dentary. Bottom, lateral occlusal view of head, showing the upper and lower tooth plates and the smaller, marginal plates supported by cartilage. In the later stages of ontogeny, the dentary tooth plate is completely resorbed. Material courtesy of J. Joss. Abbreviations: den.tp., dentary tooth plate; pre.tp., pre-articular tooth plate; pte.tp., pterygoid tooth plate; sym.t., symphyseal tooth position; vom.tp., vomerine tooth plate.

ginal tooth plates are lost. Almost the same pattern of loss occurs in *Andreivichthys* because the marginal tooth plates and the dentary bone of hatchlings subsequently disappear. The hatchling dentition is represented by dozens of dentary tooth plates, but none have been found among the thousands of juvenile and adult specimens.

Our results indicate that a specific developmental mechanism for programmed loss was preserved in these two lungfish throughout 360 Myr of evolutionary history. Early, more basal lungfish³ have palatal and lingual tooth plates, whereas marginal bones have separate teeth; some later taxa also develop and retain marginal tooth plates as adults⁵, indicating that this programme of development and loss evolved within lungfish, perhaps during the Late Devonian period.

Robert R. Reisz*, Moya M. Smith†

*University of Toronto in Mississauga,
 3359 Mississauga Road, Mississauga,
 Ontario L5L 1C6, Canada

†Dental Institute, King's College London,
 Guy's Tower, London SE1 9RT, UK
 e-mail: moya.smith@kcl.ac.uk

- Smith, M. M. *Mem. Mus. Natn. Hist.* **53**, 179–194 (1988).
- Krupina, N. I. & Reisz, R. R. *Modern Geol.* **24**, 99–108 (1999).
- Smith, M. M. & Krupina, N. I. *J. Anat.* (in the press).
- Kemp, A. J. *Morph.* **225**, 345–355 (1995).
- Krupina, N. I., Reisz, R. R. & Scott, D. *Can. J. Earth. Sci.* (in the press).

Many-body and correlation effects in semiconductors

D. S. Chemla* & Jagdeep Shah†

* Department of Physics, University of California at Berkeley, and Materials Science Division, Lawrence Berkeley Laboratory, Berkeley, California 94720, USA

† Bell Laboratories, Lucent Technologies, Holmdel, New Jersey 07733, USA

Solids consist of 10^{22} – 10^{23} particles per cubic centimetre, interacting through infinite-range Coulomb interactions. The linear response of a solid to a weak external perturbation is well described by the concept of non-interacting ‘quasiparticles’ first introduced by Landau. But interactions between quasiparticles can be substantial in dense systems. For example, studies over the past decade have shown that Coulomb correlations between quasiparticles dominate the nonlinear optical response of semiconductors, in marked contrast to the behaviour of atomic systems. These Coulomb correlations and other many-body interactions are important not only for semiconductors, but also for all condensed-matter systems.

One of the most fascinating properties of quantum mechanical systems is that they can be in entangled states, that is, a coherent superposition of eigenstates unknown in classical systems. Entangled states have been observed in atomic physics, but they remain elusive in condensed matter. The primary difficulty is that they survive only as long as the coherence is maintained, and numerous processes conspire to produce decoherence, dephasing and relaxation on an extremely short timescale ($\sim 10^{-12}$ s) in condensed matter, which is typically composed of 10^{22} – 10^{23} particles cm^{-3} that interact through the infinite-range Coulomb force. The complexity of the problem makes any theoretical approach extremely difficult. About five decades ago, Landau proposed that real particles, strongly interacting among themselves but evolving in the real vacuum, be mapped onto ‘quasiparticles’. The quasiparticles are ‘dressed’ by a part of the interaction, and are relatively long-lived excitations of the many-body system evolving in a ‘new vacuum’ that consists of the ‘rest’ of the many-body system and the part of the Coulomb interaction not accounted for in the quasiparticles. The quasiparticles are complex objects (Cooper pairs in a superconductor, excitons in a semiconductor, and so forth), and the new vacuum itself can be structured (it can have an antiferromagnetic order in the copper oxides) and dynamical (it can sustain phonons and magnons). For many solids, the concept of quasiparticles has been very useful for describing the ground state and small departures from it, that is, the most basic excitation and the linear response to weak external perturbations. However, the part of the Coulomb forces not accounted for in the formation of quasiparticles leads to interactions between these quasiparticles, inducing nonlinearities and destroying their phase coherence. Scientists interested in the linear response of a many-body system like to call these interactions ‘residual’. They are, however, not small, as shown below. They are residual only in the sense that we do not know (yet) how to treat them correctly. Understanding the influence of these many-body interactions and Coulomb correlations is one of the central challenges of condensed-matter physics.

Semiconductors form an ideal laboratory for quantitatively investigating the role of Coulomb correlations for several important reasons. Their linear properties are well understood because of a wealth of theoretical background and well established approximations that have been developed over several decades, and because of the nearly perfect samples of bulk materials and nanostructures that are now available. Furthermore, the small effective mass and large dielectric constant of semiconductors allow one to explore in ‘tabletop’ experiments many physical regimes that would require

extreme, and often unattainable, conditions in other systems. For example, the exciton Rydberg in the model semiconductor gallium arsenide (GaAs) is only 4.2 meV, about 3,000 times smaller than the hydrogen Rydberg (13.6 eV). Consequently, in GaAs the exciton ionization field, about $1 \text{ V } \mu\text{m}^{-1}$, can be easily achieved. Similarly, the magnetic field for which the cyclotron radius is equal to the Bohr radius (about 3.5 T in GaAs) is easily delivered by commercial magnets, whereas for hydrogen it is about 10^4 T, and is only found at the surface of neutron stars. Finally, semiconductors are a prototype for many condensed-matter systems so that the information about Coulomb correlations that we can obtain by studying them is valuable in understanding correlations in more complex systems.

Quasiparticles, such as excitons, dominate the linear optical properties of semiconductors near the electronic bandgap. Therefore, interactions between these quasiparticles profoundly influence the nonlinear optical response of semiconductors. In order to understand all the ubiquitous processes associated with Coulomb correlations, it is clear that one must drive the many-body systems away from their ground state and investigate the nonlinear regime. Indeed, a number of investigations during the past decade have unambiguously demonstrated that Coulomb interactions dominate the nonlinear optical response of semiconductors, in contrast to atoms, whose nonlinear response is governed solely by the Pauli exclusion principle. Significant new understanding of the physics of many-body and Coulomb correlation effects have been obtained through time-resolved nonlinear optical spectroscopy. A review of these results, and their implications, is appropriate at this time. Our goal is to emphasize the physical insights recently obtained and to discuss how they can further our understanding of Coulomb correlations in condensed matter in general.

We begin by reviewing two seminal experiments^{1–3} that clearly demonstrate the richness in the nonlinear response of semiconductors, and its fundamental and profound differences from atoms. In particular, the response of semiconductors cannot be understood in terms of the non-interacting two-level model⁴ that has worked so well for dilute atomic systems. We will show that Coulomb interaction is responsible for this difference, and discuss a mean-field formalism, the semiconductor Bloch equations⁵, developed to understand these experiments. In the stationary limit, this mean-field theory shares many features with the Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity; this shows that laser-excited semiconductors are indeed governed by the same type of interactions as other quantum-coherent many-body systems in condensed matter.

Although the mean-field theory has enjoyed considerable success

at moderate and high densities, recent studies in regimes of low excitation and very short timescale show that the behaviour of semiconductors is even more complex^{6,7}. In these regimes one can probe timescales shorter than the time between quasiparticle collisions so that there are not enough scattering events over the time span of an experiment to establish the mean-field conditions. These investigations, therefore, access regimes where the fluctuations in the scattering between elementary excitations induce large fluctuations of the mean-field order parameters. Then four-particle and higher-order correlations become important and even dominant. Many-particle correlation functions appear in almost all theoretical approaches for describing condensed matter; however, it is usually extremely difficult to access these quantities experimentally. These recent studies^{6,7} allow for the first time the direct

investigation of high-order correlations. We will discuss the new insights into many-body physics provided by such experiments.

Finally, Fermi's golden rule and Boltzmann kinetics are two cornerstones of solid-state theory and are the basis for calculating transition and scattering probabilities between two states. They assume that collisions are instantaneous and only the mean free time between the collisions matters. These two cornerstones of solid-state theory are no longer valid on timescales much shorter than the characteristic periods of the entities involved in a 'collision'. The wave-like nature of excitations in solids tells us that all interactions between waves are nothing but interference phenomena, which therefore have a finite duration on the scale of at least one period of the interacting waves. Collisions cannot be considered as 'complete' for times shorter than these periods, and the

Box 1

Principle of time-resolved nonlinear optical spectroscopy experiments

The dynamical response of a semiconductor following excitation by an ultrafast laser pulse can be conveniently divided into coherent and incoherent regimes. Timescales for various regimes depend on the nature of quasiparticles and have been discussed earlier¹⁴. Coulomb correlation effects, the focus of this review, manifest themselves most directly in the coherent response of semiconductors. Four- and six-wave-mixing experiments probe the coherent response (polarization) of the medium, pump/probe experiments investigate both the polarization and the population dynamics, and time-resolved luminescence experiments probe primarily the population or incoherent dynamics.

The generic experimental configuration for time-resolved nonlinear optics experiments is shown in the figure. One, generally weak, laser pulse ($E_1(t)$) propagating along direction $\hat{k}_1$ excites a sample at $t = 0$. Alone, this pulse would probe the linear properties of the sample, and hence it is often called the probe pulse. A second laser pulse (the pump pulse), $E_2(t)$, propagating along $\hat{k}_2$, is delayed and excites the sample at $t = \tau_d$. Because of the nonlinearities in semiconductors the response of the sample to the total field, $E_T(t) = E_1(t) + E_2(t)$ is not the sum of the responses to each field. In general, two types of measurements can be performed. They are called pump/probe and coherent wave mixing experiments^{12,14,15}.

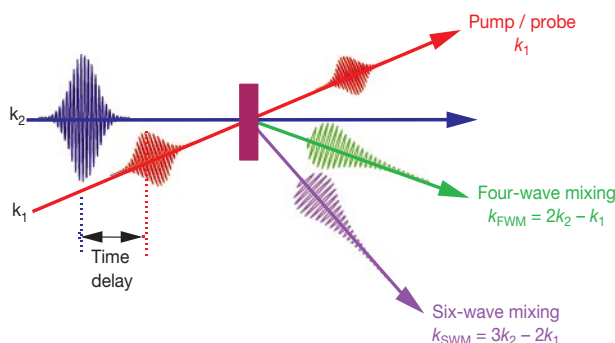
In the first category the small changes in sample transmission, seen by the probe pulse and induced by the pump pulse, are measured. In the small signal regime, the differential transmission spectrum reproduces the changes in the absorption spectrum of the sample induced by the second field. It can be measured for a series of time delays τ_d .

In the second category of experiments, the two fields coherently interfere in the sample and generate polarization waves emitting photons in background-free and momentum-conserving directions (the phase is conserved in a coherent process). For example, at the third order in the laser fields, photons are emitted in the direction $\hat{k}_{\text{FWM}} = 2\hat{k}_2 - \hat{k}_1$, whereas at the fifth order they are emitted in the

direction $\hat{k}_{\text{SWM}} = 3\hat{k}_2 - 2\hat{k}_1$ and so on. Because of the (minimum) number of photons involved in these processes, the experiments are called four-wave-mixing (FWM), six-wave-mixing (SWM) and so on. To understand the physical underpinning of the nonlinearity involved in this process, we can spectrally and/or temporally resolve the coherent wave mixing signals at fixed τ_d . Thus these techniques are called spectrally resolved FWM (or SWM) or time-resolved FWM (or SWM) and so forth. In the case of atomic-like systems the profile of the time-resolved signal is a step function followed by an exponentially decaying tail characterizing the dephasing of the associated polarization. For this reason, the easiest and most commonly used measurement technique is simply to integrate the coherent wave mixing signal with a slow detector as τ_d is varied. This determines the so-called time-integrated signal. For atomic systems, it reproduces, as a function of the time delay τ_d , the same behaviour as does the time-resolved signal as a function of the real time t .

In addition to measuring the intensity of the coherent response, it may be desirable to obtain information about the phase of the emitted signal. Spectral interferometry provides a means for accomplishing this³⁷. In this technique, the unknown emission is collinearly combined with a suitably delayed, well characterized, laser pulse whose spectrum overlaps that of the unknown emission, and the spectra of the unknown, the laser and the combined beams are measured. A spectral interferogram, obtained by subtracting the two individual spectra from the combined spectrum, exhibits spectral fringes corresponding to the time delay. Fourier transform analyses of the interferogram allow us to determine the amplitude and the phase of the unknown signal in the temporal and spectral domains. The results can be plotted in a Wigner-function representation.

In contrast to transport experiments that probe equilibrium properties averaged over many parameters, ultrafast nonlinear optical experiments allow us to probe a highly nonequilibrium regime, and provide information not possible through other means.



Box 1 Figure Schematic of a generic time-resolved nonlinear optical experiment.

formalism of quantum kinetics instead of Boltzmann kinetics must be used. We will present examples of quantum kinetic effects in semiconductors, and show that collision in progress can in fact be reversed under appropriate conditions⁸. We note that intracollisional field effects have been considered quite extensively in theoretical quantum transport studies (see, for example, ref. 9). Therefore, the insights obtained from the optical studies reviewed here are also relevant to transport phenomena¹⁰.

A large number of researchers have contributed to the advances in this field. However, this review cannot be comprehensive, and thus we limit ourselves to a broad overview of the most significant developments. The reader is also referred to a number of excellent recent review articles and books^{5,11–18}.

Nonlinear optical response of semiconductors

The interaction between quasiparticles in semiconductors is best investigated by studying their coherent response. Box 1 summarizes the experimental techniques generally used for such studies and discusses the behaviour of dilute atomic systems in such experiments¹⁴. Here we discuss two seminal experiments demonstrating that the nonlinear response of semiconductors is very different from that of a non-interacting two-level system.

In the first experiment the time-integrated four-wave mixing (FWM) response of a high quality GaAs quantum-well sample was measured as a function of the lattice temperature¹ with a time resolution of about 300 fs. Figure 1 shows the measured time-integrated signal in the $2\vec{k}_2 - \vec{k}_1$ direction as a function of time delay τ_d between the pulses. τ_d is considered positive (negative) when pulse 1 (2) precedes pulse 2 (1). Figure 1 shows that a strong signal was observed for negative time delays ($\tau_d < 0$) much larger than the time resolution of the experiment. This result is in striking

qualitative contrast to the expectation of no FWM signal for $\tau_d < 0$ (ref. 4) for non-interacting independent two-level systems, for example atoms (see Box 1). The time constant for the rise of the signal is half that for the decay (inset, Fig. 1). These observations provided the first evidence that the nonlinear response of semiconductors is very different from that of atoms.

Time-resolved four-wave mixing experiments showed further surprising results^{2,3}. For atomic systems, the time-resolved signal in the $2\vec{k}_2 - \vec{k}_1$ direction is expected to peak when pulse 2 reaches the sample, and then decay exponentially with time t . Figure 2 shows the time-resolved FWM signal in the $2\vec{k}_2 - \vec{k}_1$ direction from a high-quality GaAs quantum-well sample, with $t = 0$ corresponding to the arrival of pulse 2 at the sample. In contrast to the expectations for two-level atoms, the signal peaks at time $t = t_0$, where t_0 is extremely long (as much as 10 times the laser pulse width)². The delay decreases strongly as the dephasing rate decreases with increasing temperature (Fig. 2), but was shown to be independent of τ_d , in contrast to the behaviour expected for the photon echo signal from inhomogeneous systems¹⁴. The time-integrated signal at negative time delays, and the delayed peak in the time-resolved signals were explained in terms of local fields resulting from Coulomb interactions¹⁹. A more detailed discussion of the physical origins of these effects is given in the next section.

These two experiments led to a flurry of activity, both theoretical and experimental, aimed at understanding the nonlinear response of semiconductors. A review of the early work can be found in refs 20 and 21.

The semiconductor Bloch equations

The material parameters that enter the Maxwell equations and describe nonlinear response are the polarization and the carrier density (population). In the density matrix formalism⁴, they represent the off-diagonal and diagonal components of the density matrix, respectively. In general, they are given by the expectation

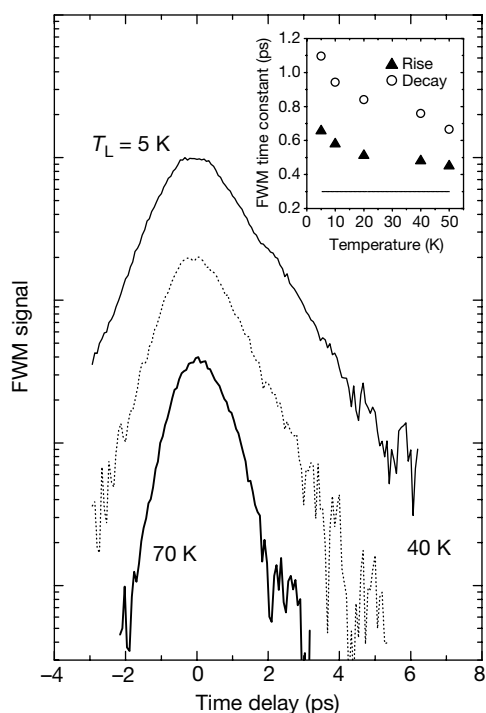


Figure 1 Strong negative time-delay signals in time-integrated four-wave mixing (FWM) from GaAs quantum wells. Time-integrated FWM signals in the $2\vec{k}_2 - \vec{k}_1$ direction versus time delay τ_d between the two pulses for a 170 Å GaAs multiple-quantum-well sample at three different temperatures. The results show an unexpected signal at negative time delays much longer than the system time resolution.

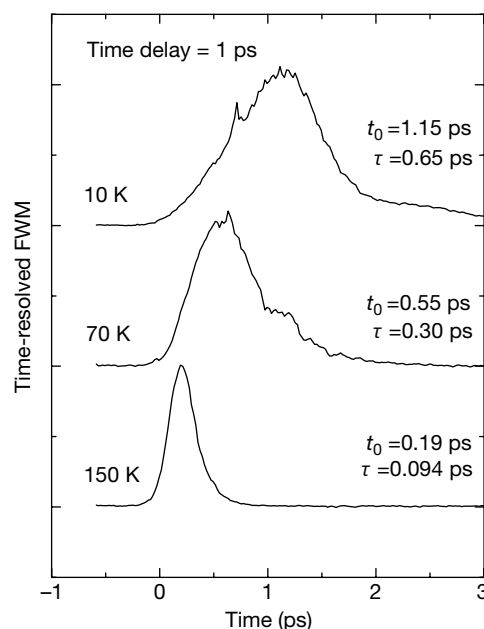


Figure 2 Delayed rise of FWM signals from GaAs quantum wells. Temporal evolution of FWM signal from a nearly homogeneously broadened 170 Å GaAs multiple-quantum-well sample showing the unexpectedly slow rise of the signal; the peak occurs at shorter times as the dephasing rate increases with increasing temperature². t_0 is defined in the text; τ is the decay time constant.

value of products of two creation and destruction operators. Semiconductors are different from an ensemble of two-level systems because Coulomb interactions between electrons and/or holes introduces products of four creation and destruction operators in the equations of motion of these products (see Box 2). This results in an infinite hierarchy of coupled differential equations for products of 2, 4, 6 ... operators that cannot be solved in general. This 'complexity disaster' appears in all many-body formalisms: the

Feynman diagrams, Green's functions and generalized density matrix approach. A consistent truncation scheme is necessary to provide a description at an appropriate level of accuracy. Because of the complexity of the problem, however (see Box 2), the input of experiment is crucial in guiding the theorists to identify the dominant processes. An approach able to account for the lowest-order correlation, first introduced to describe the excitonic optical Stark effects²², has been developed²³. It is a mean-field theory based

Box 2

Many-body effects in light-semiconductor interaction

Here we give a very basic description of the main issues of condensed-matter many-body theory. The reader interested in going beyond this (incomplete and intuitive) description can refer to refs 5, 12, 13, 15, 18 and 28–31.

The hamiltonian of a semiconductor is governed by electromagnetic interaction only. The Coulomb interaction among electrons and nuclei can be condensed in the 'effective mass approximation' in which the electronic levels are described by a few energy bands. The real electrons are mapped onto quasiparticles (unfortunately also called 'electrons') with an effective mass ($m \rightarrow m^* \ll m$) close to the band extrema, a dispersion $\epsilon(k)$ far from these special points, and a Coulomb interaction screened by the high-frequency dielectric constant (about 10–15 in most semiconductors).

Very often, because of doping with impurities or optical excitation, a relatively small number (compared to 10^{23} cm^{-3}) of electrons are present in the conduction bands and a few electrons are 'missing' in the valence bands. It has become customary to introduce yet another quasiparticle, the holes, to describe these missing electrons.

With all these caveats, the electronic part of the semiconductor hamiltonian is written in a simplified form

$$H_{\text{SC}} = H_e + H_h + H_{e-e} + H_{h-h} + H_{e-h} \quad (1)$$

with

$$H_e = \sum_{\alpha,k} \epsilon_{\alpha}(k) \hat{e}_{\alpha,k}^+ \hat{e}_{\alpha,k} \quad (2)$$

and

$$H_h = \sum_{\beta,k} \epsilon_{\beta}(k) \hat{h}_{\beta,k}^+ \hat{h}_{\beta,k} \quad (3)$$

Here $\hat{e}_{\alpha,k}^+$, $\hat{h}_{\alpha,k}^+$, $\hat{e}_{\alpha,k}$ and $\hat{h}_{\alpha,k}$ are the creation and destruction operators for electrons (e) or holes (h) with wave-vector k in the band α , whose energy $\epsilon_{\alpha}(k)$ is described by the proper effective mass expressions. Close to the band extrema these can be approximated by parabolas, $\epsilon_{\alpha}(k) = \hbar^2 k^2 / 2m_{\alpha}^*$, characterized by the effective masses m_{α}^* .

The last three terms of the hamiltonian describe, respectively, the Coulomb scattering between electrons, between holes and between electrons and holes. They have the generic form

$$H_{e-e} = \sum_{\alpha\alpha'\beta\beta',kk'} v_q \hat{e}_{\alpha,k+q}^+ \hat{e}_{\beta,k-q}^+ \hat{e}_{\alpha',k} \hat{e}_{\beta',k'} \quad (4)$$

They contain products of four creation and destruction operators and

correspond to the fundamental Feynman diagram shown in panel **a** of the figure.

In the semi-classical approximation (the electromagnetic field (EMF) is not quantized, because the intense laser pulses contain a very large number of photons) the interaction with the light field is given by the sum of Rabi frequencies

$$H_{\text{SC-EMF}} = - \sum_{\alpha\beta,k} \mu_{\alpha\beta,k} E(t) \hat{e}_{\alpha,k}^+ \hat{h}_{\beta,-k}^+ + hc \quad (5)$$

$H_{\text{SC-EMF}}$ describes the creation and destruction of electron-hole pairs by absorption and emission of photons. The coupling with the electromagnetic field, $E(t)$, is measured by the dipole matrix elements $\mu_{\alpha\beta,k}$ for interband, $\alpha \neq \beta$, or intraband, $\alpha = \beta$, transitions and is illustrated in panel **b** of the figure.

We note the analogy between panel **b** and 'half' of panel **a**. Indeed, they both describe the scattering of particles from one state to another. The main difference, however, is that the photons correspond to the applied transverse field $E(t)$, whereas the Coulomb potential describes the mutual longitudinal field exercised by the particles on one another.

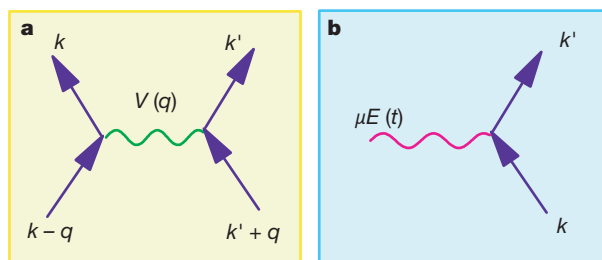
Fermion operators always appear in pairs. So when we try to describe the time evolution of the product of an even number ($2n$) of such operators under a hamiltonian containing H_{SC} we obtain equations involving at least $(2n + 2)$ operators, and that triggers the generation of an infinite number of coupled equations. This process can take different embodiments depending on the formalism used (Feynman diagram, Green's function, density matrix), but the root cause is always the same. We note that the correlation functions that appear in the theory describe properties of the electronic system as well as the scattering processes that can happen in the interacting many-body system. For example, the electron and hole populations

$$n_{e,\alpha}(k) = \langle \hat{e}_{\alpha,k}^+ \hat{e}_{\alpha,k} \rangle \quad \text{and} \quad n_{h,\beta}(k) = \langle \hat{h}_{\beta,k}^+ \hat{h}_{\beta,k} \rangle \quad (6)$$

and the transition amplitudes

$$p_{\alpha\beta}(k) = \langle \hat{e}_{\alpha,k}^+ \hat{h}_{\beta,-k}^+ \rangle \quad (7)$$

are 2-particle correlation functions: respectively, the diagonal and off-diagonal elements of the density matrix. Higher-order correlation functions often have an intuitive interpretation; in the context of this review, for example, the 4-particle correlation function, $\langle \hat{e}_{\alpha,k}^+ \hat{h}_{\beta,-k}^+ \hat{e}_{\alpha',k} \hat{h}_{\beta',-k'} \rangle$, describes the creation of two electron-hole pairs in a two-photon transition.



Box 2 Figure Feynman diagrams for (a) Coulomb interaction between carriers, and (b) coupling with electromagnetic field.

on the random phase approximation (RPA)²⁴, which is founded on the physical argument that under suitable circumstances a sum of exponential terms with randomly varying phases may be neglected. (RPA goes by many names, including time-dependent Hartree–Fock approximation.) RPA results in the factorization of the expectation value of the four-particle operators into the products of 2-particle operators. Within this theory, the time evolution of the polarization and density form a closed system, the semiconductor Bloch equations (SBE)⁵. We use this formalism here because the SBE are the semiconductor analogue of the optical Bloch equations (OBE) that are extensively used in atomic and molecular physics. The OBE are themselves the optical analogue of the equations first developed by Bloch²⁵ for the spin systems.

The details of the two seminal experiments discussed in the previous section can be described accurately by numerical solution of the SBE. However, the underlying physics can be qualitatively understood by examining the form of these equations. The SBE are directly analogous to the OBE except that the effective Rabi frequency (describing the light–matter coupling) and the energy of the quasiparticles (for example, excitons) are renormalized (that is, modified) by polarization and population respectively. Very intuitively these ‘renormalizations’ can be explained by saying that the charges in the medium feel not only the applied laser field, but also the field of all the other dipoles induced in the semiconductor. Similarly, their energy is modified by the Coulomb interaction with the other induced charges. These renormalizations profoundly influence the nonlinear response of semiconductors^{15,26,27}. For non-interacting two-level systems, the FWM signal appears as an

interaction between the field of the second pulse with the polarization created by the first pulse and thus does not exist when they do not overlap ($\tau_d < 0$). In a semiconductor, the renormalization of the Rabi frequency allows interaction between the polarizations created by the two pulses. Because polarizations persist for times of the order of the dephasing time after the exciting electric field, they overlap regardless of the pulse sequence and produce the $\tau_d < 0$ signal. Furthermore, these renormalizations allow diffraction of polarization. The third-order polarization resulting from diffraction of polarization increases with time and reaches a peak at a time determined by the dephasing time. Therefore, the time-resolved signal exhibits a delayed peak after the second pulse, as observed experimentally.

Recent results have shown that the nonlinear response of semiconductors is even more complex, and higher-order, 4-particle and 6-particle correlations make important and sometimes dominant contributions to the nonlinear response. These are discussed in the next section.

High-order correlation effects

Although the mean-field theory (SBE) has been extremely successful at high and moderate densities, at low densities we can investigate timescales that are short compared to the time for one quasiparticle to interact with a substantial fraction of its neighbours. Thus it becomes possible to observe deviations from mean-field theory, a regime where high-order correlations become dominant. Indeed, recent comparative experimental and theoretical studies have revealed features related to genuine 4-particle and 6-particle correlations. Different ways of extending the theory beyond the 2-particle, mean-field approach of SBE have been developed^{12,17,18,28–31}. Some experimental investigations of higher-order correlations^{6,7,32–34} are discussed below.

The effects of 4-particle correlations in the continuum of exciton scattering states are shown in Fig. 3 where time-integrated FWM measured in a GaAs sample with heavy-hole and light-hole exciton resonances in the absorption spectrum is shown for magnetic field varying from 0 to 10 T (ref. 6). The signal has an expected profile for $\tau_d > 0$: beats between the two exciton resonances are superimposed on an exponential decay, with only slight dependence on B . In contrast, the Coulomb-induced signal for $\tau_d < 0$ changes drastically as B increases. Not only does its magnitude increase, but also its profile becomes highly non-exponential with an unusual positive curvature and extends as far as $\tau_d \approx 10$ ps, or 100 times the duration of the laser pulse (compare with Fig. 1). The explanation of this effect is that the excitons created in the semiconductor by one laser pulse are very strongly distorted by the magnetic field. Therefore, they acquire a quadrupole moment and through the Coulomb interaction they generate a coherent 4-particle correlation in the medium. It corresponds to a two-photon active coherence³⁵ and, therefore, cannot deliver one photon emission by itself. Thus it is stored in the medium until the second pulse triggers that emission. This ‘coherent memory’ can be interpreted as a non-markovian process involving the 2-particle correlation polarization waves interacting with a bath of 4-particle correlations³⁶.

The effects of 4-particle correlations have been shown to be important in other recent experiments as well⁷. At low densities, incident radiation resonant with excitons in quantum wells is scattered (resonant Rayleigh scattering) due to disorder (and hence inhomogeneous broadening of excitons) at low exciton densities. However, 4-particle correlations can lead to intrinsic resonant Rayleigh scattering even in the absence of disorder. Spectral interferometry^{37,38} allows determination of the temporal dynamics and the spectra of the amplitude and the phase of coherent radiation. Four-particle correlations can be thought of as the bound and unbound (continuum) states of biexcitons (an entity composed of two excitons). New spectral features related to 4-particle correlations have been identified recently in resonant

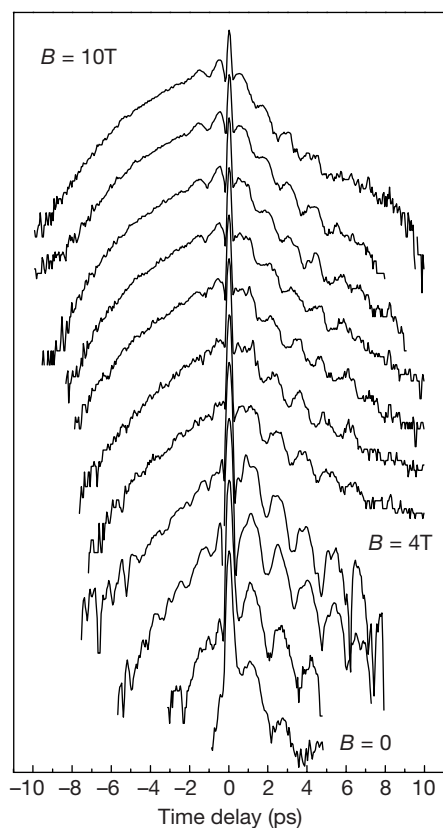


Figure 3 FWM signals due to high-order correlation in GaAs. Time-integrated FWM signal from bulk GaAs (log scale) at 1.8 K for various magnetic fields. The curves are displaced vertically for clarity. We note the huge increase in the negative time-delay component with increasing B (ref. 39).

Rayleigh scattering from high-quality GaAs quantum wells³⁸. The Wigner function provides a convenient means of displaying emitted field. The coherent spectrum and dynamics can be deduced from such a representation (Fig. 4). The coherent spectrum deduced from such measurements shows not only the disorder-induced resonant Rayleigh scattering from excitons (near 807 nm) but also intrinsic Rayleigh scattering from the bound state of the biexciton (near 808 nm) and the continuum 4-particle states (<807 nm). The dip in the response at a wavelength slightly below 807 nm is due to interference between the response from excitons and 4-particle correlations. These experiments provide the first information on the spectrum of 4-particle correlations, and have the potential of providing new understanding of the nature of 4-particle correlations.

Higher-order correlations such as six-particle correlation effects are difficult to identify in conventional FWM and pump/probe experiments. However, six-wave-mixing (SWM) experiments are expected to be more sensitive to the high-order correlations, although 2-particle and 4-particle correlations also contribute to the SWM signals. Indeed, recent comparative experimental and

theoretical studies³² have revealed genuine 6-particle correlations in a single ZnSe quantum well. Figure 5a presents the contour maps of the spectrally resolved SWM emission measured in the direction $3\vec{k}_2 - 2\vec{k}_1$ for incident laser pulses polarized along the x and y axes, and the signal polarized along the y axis. These plots give information both on the dependence of the signal at a given frequency on the time delay τ_d , and dependence on real time at a given τ_d (related to the Fourier transform of the frequency dependence at a given τ_d). Two contributions are seen at the exciton and the exciton–biexciton energies, both decaying smoothly without temporal or spectral beats. Figure 5b–d shows theoretical modelling of the experiments at three levels of sophistication. Figure 5b accounts only for the 2-particle and 4-particle correlation involved in the purely coherent emission, that is, 1-photon and 2-photon coherence. Figure 5c accounts in addition for incoherent exciton densities and Fig. 5d also accounts for 6-point correlations that describe transitions from incoherent exciton densities to two-pair states; these are often referred to as ‘excited-state absorption’ in atomic physics. The theory plots are obtained by exactly solving numerically the coupled

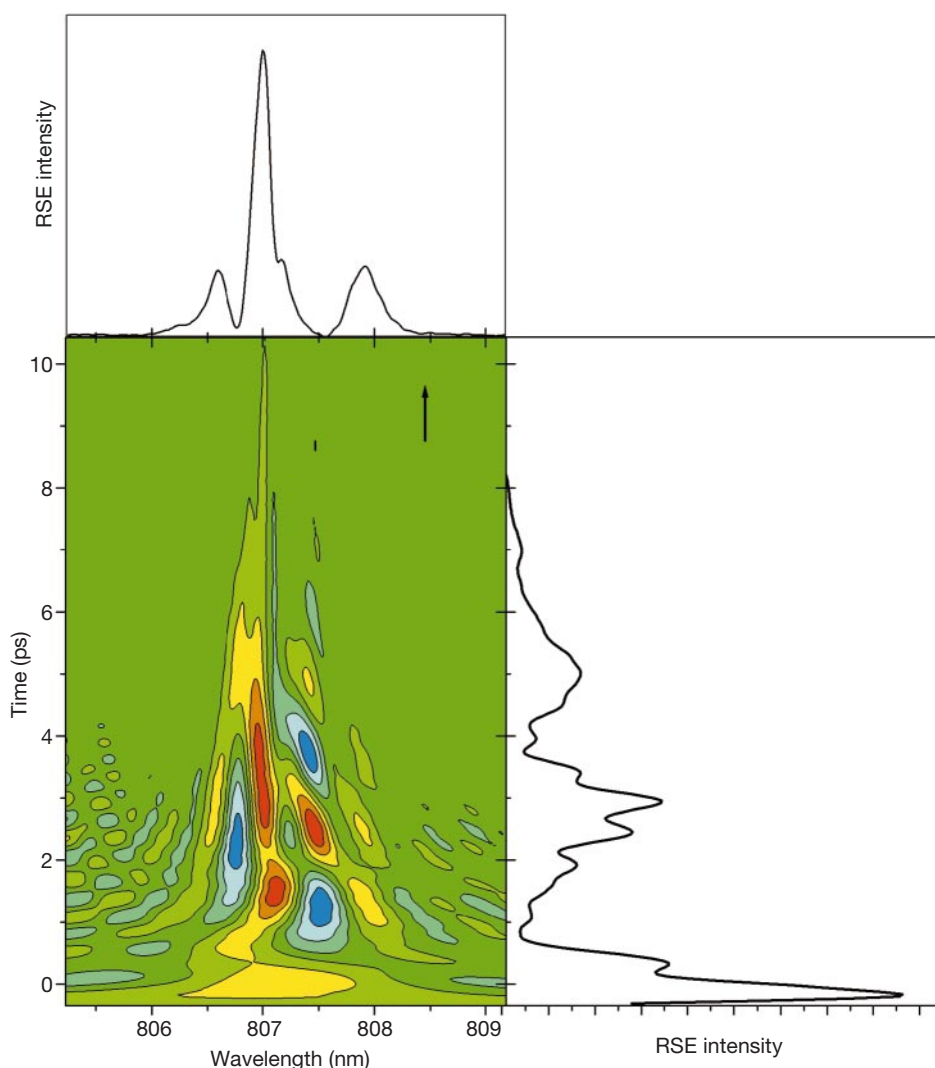


Figure 4 Wigner-function representation of resonant Rayleigh scattering from GaAs quantum wells. Wigner-function representation of the intrinsic and disorder-induced resonant Rayleigh scattering signal from a 130 Å GaAs multiple quantum-well sample at 10 K at intermediate excitation density. The coherent spectrum and the coherent dynamics can be determined from the Wigner function, as shown. The coherent spectrum shows features related to the spectral response of 4-particle correlations⁷. RSE, resonant secondary emission.

equations of motion of the 2-, 4- and 6-particle correlation functions and thus the results include the contributions of all orders in the exciting laser field. Figure 5b and c show large disagreements with experiment. But adding 6-point correlations (Fig. 5d) brings the theory into good qualitative agreement with experiment, showing that correlation functions at least of the sixth order are absolutely necessary to describe accurately the SWM experiments.

Although high-order correlation functions appear very often in the theory of many-body effects in condensed matter, they are rarely accessible to direct measurement. The experiments of the type described above provide a means of observing the effects of higher-order correlations and investigating their properties. For example, the dephasing of 4-particle correlation functions by phonons and carrier scattering has been studied recently for the first time³⁹.

Coherent control, memory effects and relaxation

Another regime where traditional assumptions fail is that of strongly correlated dynamics where it is no longer justified to treat the interaction between quasiparticles as scattering events local in space and time (a key assumption of the Boltzmann theory). In that case memory effects become observable. In principle, for any quantum mechanical system described by a Schrödinger equation, the knowledge of the hamiltonian and the present state is enough to determine the future evolution. In practice, however, we do not know how to solve the time-dependent Schrödinger equation of complex many-body systems. So it is customary to divide the total system into a subsystem that we are interested in, and the thermal ‘bath’ of all the other degrees of freedom (reservoir) on which we have only partial information. Memory effects and dissipation occur through the following mechanism: at a time t' an interaction occurs between the subsystem and some degrees of freedom of the reservoir, which then evolve with their own dynamics and at a later time, t , interact again with the subsystem. From the viewpoint of the evolution of the subsystem, processes depending on the time difference ($t - t'$) are occurring and hence the subsystem develops a ‘memory’. Because the reservoir is so large and contains so many degrees of freedom, in general there are many complex interferences between t and t' so that information is lost

and, again from the viewpoint of the subsystem, there is ‘dissipation’. However, on a timescale that is short compared to the periods of the reservoir excitation and before the reservoir/subsystem scattering processes have randomized the degrees of freedom, the details of the evolution of the coupled system (subsystem plus reservoir) must be accounted for and the dynamics of the subsystem must be described by quantum kinetics, with non-markovian statistics as we mentioned in the introduction. The main relaxation mechanisms in semiconductors are electron–electron and phonon–electron scattering. The timescale of the initial phonon–electron non-markovian regime can be estimated by noting that the time it takes a lattice to react to a perturbation is of the order of one period of its oscillation, that is, in the case of the longitudinal optical (LO) phonon in GaAs, $T_{LO} = 115$ fs.

Memory effects in the nonlinear response of semiconductors have indeed been observed^{8,34,40–44}. Here we first discuss non-markovian effects in the coupling of an LO phonon with interband polarization observed in bulk GaAs by FWM experiments using pulses of about 14 fs, about eight times shorter than the LO-phonon period⁸. Time-integrated FWM exhibits an oscillatory modulation superimposed on the usual exponential decoherence decay (Fig. 6; curves 7 to 10). The period, about 98 fs, corresponds to the separation between conduction band states coupled by one LO-phonon. These oscillations are not due to emission or absorption of phonons but correspond to a coherent memory stored in the LO-phonon bath that modulates the interband transition. This phonon–electron interaction, however, is not ‘irreversible’. Like any other quantum mechanical ‘collision’ it is an interference that can be reversed or enhanced by a second perturbation whose phase is appropriately chosen. This reversal was indeed experimentally demonstrated, as shown in Fig. 6. The detailed theory of these experiments is not yet available, but they demonstrate that by using coherent control of the excitation it is possible to manipulate the so-called ‘irreversible’ scattering processes.

In some special systems it is known from transport experiments that electron–electron interaction can be very significantly reduced. For example, under high magnetic field a two-dimensional electron gas (2DEG) enters the regime of the quantum Hall effect (QHE) and at integer filling factors behaves as an incompressible fluid with a

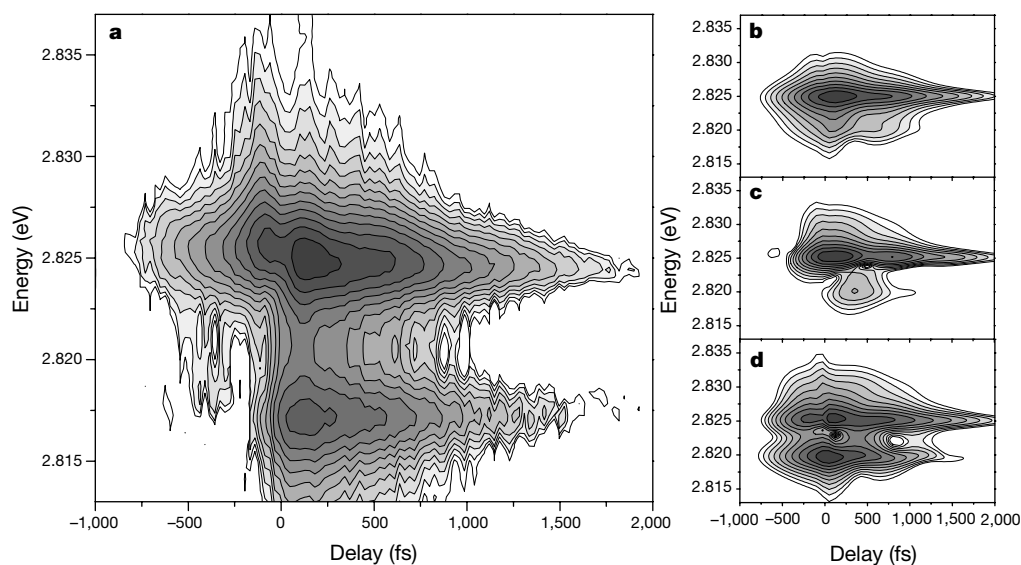


Figure 5 Signature of higher-order correlations in the coherent response of ZnSe. Contour map of spectrally resolved six-wave-mixing (SWM) signal as a function of time delay for a ZnSe quantum well. **a**, Experiment, with greyscale representing variation of the SWM signal over three orders of magnitude; **b–d**, calculated SWM for three different cases discussed in the text³².

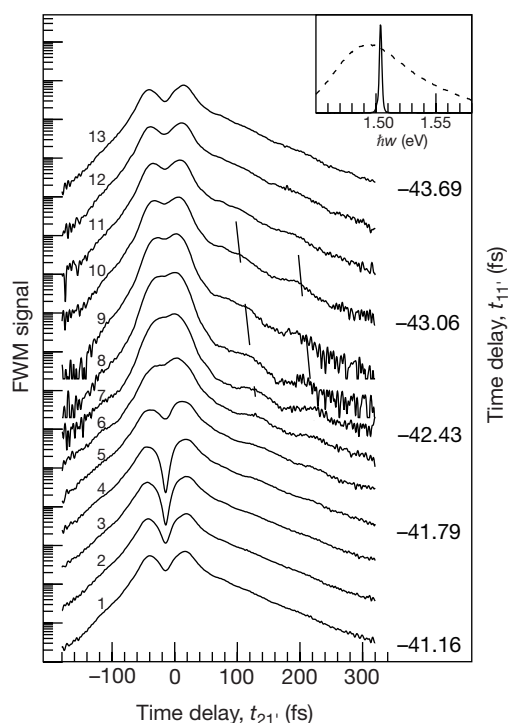


Figure 6 Reversible electron-photon scattering in GaAs. Time-integrated FWM signal from bulk GaAs at 77 K obtained by using two phase-locked pulses (1 and 1') along $\vec{k}_1$ and a single pulse (2) along $\vec{k}_2$. The FWM signals are displayed as a function of time-delay t_{21} , for many different delays t_{11} . The curves are displayed vertically for clarity. The inset shows the exciton and laser spectra⁸.

phase coherence time that diverges at low temperature. One thus expects that the quantum coherence of the interband polarization will exhibit peculiar properties under these conditions. Such behaviour has been observed recently⁴⁵ in FWM experiments performed on 2DEG in the QHE regime, where the laser pulses were tuned to excite electrons into only the highest partially occupied Landau level containing the Fermi energy. At high magnetic field when the Landau levels are only fractionally filled, the time-integrated FWM signal has a complicated profile with a non-exponential behaviour and a long decay time for short time delays. This profile evolves to a single exponential with an unusually long decay time when the Landau level approaches full filling, but the dephasing time experiences a very sudden decrease when the Fermi energy moves to the next Landau level. A measure of the interband polarization dephasing time is obtained by extracting an overall decay time from the FWM profiles and is displayed in Fig. 7 (squares) for two samples with different doping. We note the very large jump each time the system passes through even filling factors for samples with very different densities, showing that this is an effect of the cold 2DEG. When the Landau levels are partially filled, the dephasing originates mainly from the scattering of the photoexcited carriers with the intra-Landau-level collective excitations of the strongly correlated two-dimensional electron liquid. A model based on scattering with the 2DEG magneto-rotons⁴⁸ accounts qualitatively for the salient features seen experimentally as shown in Fig. 7 (circles).

Future directions

The dominant role of Coulomb correlations in the nonlinear response of semiconductors is now well established. Understanding higher-order correlations in semiconductors remains a challenge

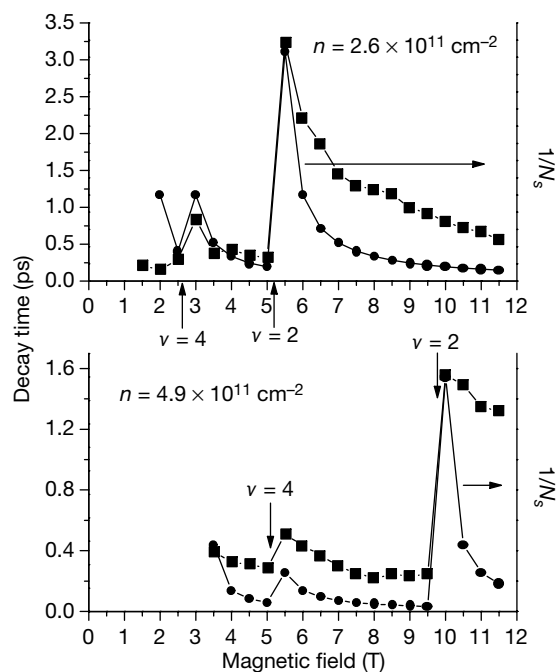


Figure 7 Coherent response of GaAs in the quantum Hall regime. Decay time (squares) of the time-integrated FWM signals from the electron gas in two modulation-doped GaAs quantum-well samples in the fractional quantum Hall regime at 2 K as a function of the magnetic field that changes spacings between Landau levels and hence the fractional occupancy of Landau levels. The circles correspond to $1/N_s = \nu/[2(N+1)-\nu]$, where N is the Landau level number and ν is the filling factor⁴⁵.

and is a good direction for future investigations. We believe that an even more important challenge is to relate this improved understanding of Coulomb correlations in semiconductors to strongly correlated solids, and investigate whether some of the approaches that have been so fruitful in semiconductors may lead to a better understanding of these solids. High- T_c superconductors, transition metal oxides and polymers are some examples of highly correlated systems. Some work in this direction has already begun^{46,47}. □

1. Leo, K. *et al.* Effects of coherent polarization interactions on time-resolved degenerate four-wave mixing. *Phys. Rev. Lett.* **65**, 1340–1343 (1990).
2. Kim, D. S. *et al.* Unusually slow temporal evolution of femtosecond four-wave-mixing signals in intrinsic GaAs quantum wells: direct evidence for the dominance of interaction effects. *Phys. Rev. Lett.* **69**, 2725–2728 (1992).
3. Weiss, S., Mycek, M.-A., Bigot, J.-Y., Schmitt-Rink, S. & Chemla, D. S. Collective effects in excitonic free induction decay: do semiconductors and atoms emit coherent light in different ways? *Phys. Rev. Lett.* **69**, 2685–2689 (1992).
4. Allen, L. & Eberly, J. H. *Optical Resonance and Two-Level Atoms* (Wiley Interscience, New York, 1975).
5. Haug, H. & Koch, S. W. *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (World Scientific, Singapore, 1993).
6. Kner, P., Bar-Ad, S., Marquezini, M. V., Chemla, D. S. & Schäfer, W. Magnetically enhanced exciton-exciton correlations in semiconductors. *Phys. Rev. Lett.* **78**, 1319–1322 (1997).
7. Birkedal, D., Shah, J. & Pfeiffer, L. N. Coherent secondary emission from resonantly excited two-exciton states: An intrinsic phenomenon. *Phys. Rev. B* **60**, 15585–15588 (1999).
8. Wehner, M. U., Ulm, M. H., Chemla, D. S. & Wegener, M. Coherent control of electron-LO-phonon scattering in bulk GaAs. *Phys. Rev. Lett.* **80**, 1992–1995 (1998).
9. Rossi, F., Brunetti, R. & Jacoboni, C. in *Hot Carriers in Semiconductor Nanostructures: Physics and Applications* (ed. Shah, J.) 153–188 (Academic, Boston, 1992).
10. Haug, H. & Jauho, A. P. *Quantum Kinetics in Transport and Optics of Semiconductors* (ed. Klitzing, K. V.) (Springer Series in Solid-State Sciences, Springer, Berlin, 1996).
11. Steel, D. G., Wang, H., Jiang, M., Ferrio, K. B. & Cundiff, S. T. in *Coherent Optical Interactions in Solids* (ed. Phillips, R. T.) 157–180 (Plenum, New York, 1994).
12. Mukamel, S. *Principles of Nonlinear Optical Spectroscopy* (Oxford Univ. Press, New York, 1995).
13. Binder, R. & Koch, S. W. Nonequilibrium semiconductor dynamics. *Prog. Quant. Electron.* **19**, 307–462 (1995).
14. Shah, J. *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* (Springer, Heidelberg, 1999).

15. Chemla, D. S. in *Nonlinear Optics in Semiconductors* (ed. Garmire, E. & Kost, A. R.) 175–256 (Academic, New York, 1999).
16. Khitrova, G., Gibbs, H. M., Jahnke, F., Kira, M. & Koch, S. W. Nonlinear optics of normal-mode-coupling in semiconductor microcavities. *Rev. Mod. Phys.* **71**, 1591–1639 (1999).
17. Perakis, I. E. & Shahbazyan, T. V. Canonical transformation approach to the ultrafast nonlinear optical dynamics of semiconductors. *Int. J. Mod. Phys. B* **13**, 869–893 (1999).
18. Perakis, I. E. & Shahbazyan, T. V. Many-body correlation effects in the ultrafast non-linear optical response of confined Fermi seas. *Surf. Sci. Rep.* **40**, 3–74 (2000).
19. Wegener, M., Chemla, D. S., Schmitt-Rink, S. & Schäfer, W. Line shape of time-resolved four-wave mixing. *Phys. Rev. A* **42**, 5675–5683 (1990).
20. Henneberger, F., Schmitt-Rink, S. & Göbel, E. O. (eds) *Optics of Semiconductor Nanostructures* (Akademie, Berlin, 1993).
21. Phillips, R. T. (ed.) *Coherent Optical Processes in Semiconductors* (Plenum, New York, 1994).
22. Schmitt-Rink, S. & Chemla, D. S. Collective excitations and the dynamical Stark effect in a coherently driven exciton system. *Phys. Rev. Lett.* **57**, 2752–2755 (1986).
23. Lindberg, M. & Koch, S. W. Effective Bloch equations for semiconductors. *Phys. Rev. B* **38**, 3342–3350 (1988).
24. Pines, D. & Nozieres, P. *The Theory of Quantum Liquids* (Benjamin, New York, 1966).
25. Bloch, F. Nuclear induction. *Phys. Rev.* **70**, 460–474 (1946).
26. Schmitt-Rink, S., Mukamel, S., Leo, K., Shah, J. & Chemla, D. S. Stochastic theory of time-resolved four-wave mixing in interacting media. *Phys. Rev. A* **44**, 2124–2129 (1991).
27. Schäfer, W., Jahnke, F. & Schmitt-Rink, S. Many-particle effects on transient four-wave-mixing signals in semiconductors. *Phys. Rev. B* **47**, 1217–1220 (1993).
28. Axt, V. M. & Stahl, A. A dynamics-controlled truncation scheme for the hierarchy of density matrices in semiconductor optics. *Z. Phys. B* **93**, 195–204 (1994).
29. Maialle, M. Z. & Sham, L. J. Interacting electron theory of coherent nonlinear response. *Phys. Rev. Lett.* **73**, 3310–3313 (1994).
30. Östreich, T., Schönhammer, K. & Sham, L. J. Exciton–exciton correlation in the nonlinear optical regime. *Phys. Rev. Lett.* **74**, 4698–4701 (1995).
31. Schäfer, W., Löwenich, R., Fromer, N. A. & Chemla, D. S. From coherently excited highly correlated states to incoherent relaxation processes in semiconductors. *Phys. Rev. Lett.* **86**, 344–347 (2001).
32. Bolton, S. R., Neukirch, U., Sham, L. J., Chemla, D. S. & Axt, V. M. Demonstration of sixth-order Coulomb correlations in a semiconductor single quantum well. *Phys. Rev. Lett.* **85**, 2002–2005 (2000).
33. Meier, T., von Plessen, G., Thomas, P. & Koch, S. W. Coherent electric-field effects in semiconductors. *Phys. Rev. Lett.* **73**, 902–905 (1994).
34. Haring Bolivar, P. *et al.* Excitonic emission of THz radiation: experimental evidence of the shortcomings of the Bloch equation method. *Phys. Rev. Lett.* **78**, 2232–2235 (1997).
35. Cohen-Tannoudji, C., Dupont-Roc, J. & Grynberg, G. *Atom–Photon Interactions: Basic Processes and Applications* (Wiley Interscience, New York, 1992).
36. Kner, P. *et al.* Effect of magnetoexciton correlations on the coherent emission of semiconductors. *Phys. Rev. B* **60**, 4731–4748 (1999).
37. Lepetit, L., Cheriaux, G. & Joffe, M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **12**, 2467–2474 (1995).
38. Birkedal, D. & Shah, J. Femtosecond spectral interferometry of resonant secondary emission from quantum wells: resonance Rayleigh scattering in the nonergodic regime. *Phys. Rev. Lett.* **81**, 2372–2375 (1998).
39. Kner, P., Schafer, W., Löwenich, R. & Chemla, D. S. Coherence of four-particle correlations in semiconductors. *Phys. Rev. Lett.* **81**, 5386–5389 (1998).
40. Banyai, L. *et al.* Exciton-LO-phonon quantum kinetics: evidence of memory effects in bulk GaAs. *Phys. Rev. Lett.* **75**, 2188–2191 (1995).
41. Bar-Ad, S., Kner, P., Marquezini, M. V., Chemla, D. S. & El Sayed, K. Carrier dynamics in the quantum kinetic regime. *Phys. Rev. Lett.* **77**, 3177–3180 (1996).
42. Camescasse, F. X. *et al.* Ultrafast electron redistribution through Coulomb scattering in undoped GaAs: experiment and theory. *Phys. Rev. Lett.* **77**, 5429–5432 (1996).
43. Bar-Ad, S. & Chemla, D. S. Quantum kinetics regime during and immediately after laser excitation of semiconductors. *Mater. Sci. Eng. B* **48**, 83–87 (1997).
44. Fürst, C., Leitenstorfer, A., Laubereau, A. & Zimmermann, R. Quantum kinetic electron–phonon interaction in GaAs: energy nonconserving scattering events and memory effects. *Phys. Rev. Lett.* **78**, 3733–3736 (1997).
45. Fromer, N. A. *et al.* Electronic dephasing in the quantum Hall regime. *Phys. Rev. Lett.* **83**, 4646–4649 (1999).
46. Dodge, J. S. *et al.* Time-resolved optical observation of spin-wave dynamics. *Phys. Rev. Lett.* **83**, 4650–4653 (1999).
47. Kaindl, R. A. *et al.* Ultrafast mid-infrared response of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Science* **287**, 470–473 (2000).
48. Garvin, S. M., McDonald, A. H. & Platzman, P. M. Magneto-roton theory of collective excitation in the fractional quantum Hall effect. *Phys. Rev. B* **33**, 2481–2494 (1986).

Acknowledgements

The work of D.S.C. was supported by the Director, Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences and Office of Science, US Department of Energy.

Correspondence and requests for materials should be addressed to J.S. (e-mail: jags@lucent.com).

High critical current density and enhanced irreversibility field in superconducting MgB₂ thin films

C. B. Eom[†], M. K. Lee^{*}, J. H. Choi[†], L. J. Belenky^{*}, X. Song[†], L. D. Cooley[†], M. T. Naus[†], S. Patnaik[†], J. Jiang[†], M. Rikel[†], A. Polyanskii[†], A. Gurevich[†], X. Y. Cai[†], S. D. Bu^{*}, S. E. Babcock[†], E. E. Hellstrom[†], D. C. Larbalestier[†], N. Rogado[‡], K. A. Regan[‡], M. A. Hayward[‡], T. He[‡], J. S. Slusky[‡], K. Inumaru[‡], M. K. Haas[‡] & R. J. Cava[‡]

^{*} Department of Materials Science and Engineering, 1509 University Avenue; and

[†] Applied Superconductivity Center, 1500 Engineering Drive, University of Wisconsin, Madison, Wisconsin 53706, USA

[‡] Department of Chemistry and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544, USA

The discovery of superconductivity at 39 K in magnesium diboride¹ offers the possibility of a new class of low-cost, high-performance superconducting materials for magnets and electronic applications. This compound has twice the transition temperature of Nb₃Sn and four times that of Nb-Ti alloy, and the vital prerequisite of strongly linked current flow has already been demonstrated^{2–5}. One possible drawback, however, is that the magnetic field at which superconductivity is destroyed is modest. Furthermore, the field which limits the range of practical applications—the irreversibility field $H^*(T)$ —is approximately 7 T at liquid helium temperature (4.2 K), significantly lower than about 10 T for Nb-Ti (ref. 6) and ~20 T for Nb₃Sn (ref. 7). Here we show that MgB₂ thin films that are alloyed with oxygen can exhibit a much steeper temperature dependence of $H^*(T)$ than is observed in bulk materials, yielding an H^* value at 4.2 K greater than 14 T. In addition, very high critical current densities at 4.2 K are achieved: 1 MA cm^{–2} at 1 T and 10⁵ A cm^{–2} at 10 T. These results demonstrate that MgB₂ has potential for high-field superconducting applications.

Three MgB₂ films were prepared by pulsed laser deposition at room temperature onto (111) oriented single crystal SrTiO₃ substrates. The base pressure before deposition was 3×10^{-6} Pa, and deposition took place under 0.3 Pa of argon. The MgB₂ target was prepared by pressing and sintering, as described previously^{2,8}. After deposition, film 1 was annealed in an evacuated niobium tube at 850 °C for 15 minutes. Film 2 was annealed in an evacuated quartz tube at 750 °C for 30 minutes and quenched to room temperature. Film 3 was annealed in a tantalum envelope inside an evacuated niobium tube at 750 °C for 30 minutes. Magnesium pellets were included in each tube to prevent magnesium loss. Film 1 showed visible evidence of inhomogeneity in the centre of the film, which could not be excluded from samples used for electromagnetic measurements. The thickness of each film was approximately 500 nm.

The chemical compositions of the films were determined by wavelength dispersive X-ray spectroscopy using an electron microprobe at 3 keV to avoid X-ray generation in the substrate. Before annealing, the films had a Mg:B:O atomic ratio of about 1.0:1.0:0.17, the oxygen probably coming from the target or background gas during deposition. The primary differences in the films are the measured ratios of Mg:B:O obtained after annealing. Film 2 had a ratio of ~1.0:0.9:0.7, which corresponds to a relative oxygen concentration approximately twice that of film 1 and 3, as well as higher Mg content. By contrast, film 1 and film 3 show 1.0:1.0:0.4 and 1.0:1.2:0.3 ratios, respectively. We believe that the higher oxygen of post-annealed film 2 came from the relatively poor vacuum (approximately 150 Pa) and the larger volume of the tube used for annealing. X-ray diffraction also revealed that the MgB₂ lattice parameter c is larger in film 2 than in film 1, film 3, or bulk samples.

Figure 1 presents θ – 2θ scans for films 1, 2 and 3, which clearly show that the (002) MgB₂ peak of film 2 is shifted to lower angle. We believe the expansion of the c -lattice parameter in MgB₂ is due to oxygen alloying of the boron layer.

The transition temperature T_c was determined by both inductive and resistive measurements. The resistive transitions, shown in Fig. 2, have onset values of 36, 31 and 35 K for films 1, 2 and 3 respectively, and fall to zero resistance just above the principal inductive transitions. The room temperature resistivity $\rho(300\text{ K})$ of film 2, 390 $\mu\Omega\text{ cm}$, is clearly much higher than that of film 1 (125 $\mu\Omega\text{ cm}$) or film 3 (60 $\mu\Omega\text{ cm}$). High-resistivity film 2 also exhibits a slight resistivity increase with decreasing temperature, indicating a resistivity ratio, $R_R = \rho(300\text{ K})/\rho(40\text{ K})$, less than 1. By contrast, the R_R of film 1 is about 1.5, whereas that of film 3 is 1.3. For comparison, the R_R of Nb-Ti alloys (Ginzburg–Landau parameter $\kappa \approx 40$) is seldom higher than 1.4 (ref. 6), whereas that of Nb₃Sn ($\kappa \approx 30$) is typically between 2 and 5 (ref. 7). Good bulk samples of MgB₂ can attain R_R values of up to 25 and have $\kappa \approx 25$ (refs 4 and 5).

The magnetization of 3 mm $\times$ 3 mm pieces of each film was measured from 4.2 to 30 K and in perpendicular fields up to 14 T using a vibrating sample magnetometer (VSM). Hysteresis curves at 4.2 K (Fig. 3a) close at fields very much higher for film 2, at about 14 T, than for film 1 (~6 T) or film 3 (~6.5 T). We were not able to unambiguously determine H_{c2} owing to the smooth blending of the superconducting magnetization into the background magnetization of substrate and sample holder. This is in marked contrast to the behaviour seen for bulk samples; these have a significant reversible contribution to the remaining hysteretic magnetic moment above the primary loop closure field^{2,4,9–11}. The oxygen-rich film 2 has a much higher irreversibility field than films 1 and 3 up to 25 K.

Magneto-optical examination of the films showed strong flux shielding with only slightly distorted roof patterns, indicative of macroscopic supercurrents. This validates use of the standard expression¹² for the magnetization of a thin film, $J_c = 30\Delta m V^{-1} r^{-1}$, to extract critical current densities $J_c(H, T)$, where Δm is the width of the hysteresis loop, $V \approx 4.5 \times 10^{-6}\text{ cm}^3$ is the film volume, and $r = 0.15\text{ cm}$ is the sample half-width. Figure 3b shows that the highest J_c value at 1 T, 4.2 K exceeds 3 MA cm^{–2} for the low-resistivity film 3. This result is comparable to that of ref. 13.

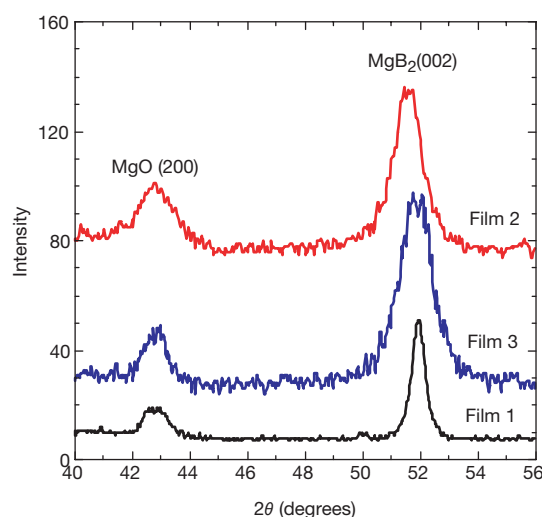


Figure 1 X-ray diffraction θ – 2θ scans of films 1 (black curve), 2 (red curve) and 3 (blue curve), showing both (002) MgB₂ and (002) MgO peaks. The 2θ position of the MgB₂ (002) peak for film 2 (51.5°) is smaller than for films 1 and 3 (51.9°). The lattice parameter c of MgB₂ in film 2 is $0.3547 \pm 0.0007\text{ nm}$, substantially larger than the 0.3521 nm values for films 1 and 3 and also for bulk material^{1,3}. The sample was tilted at about 0.5° to avoid the strong (111) SrTiO₃ substrate diffraction.

However, the much larger H^* (4.2 K) of film 2 enables J_c to achieve 1 MA cm^{-2} at 1 T and to remain above 10^5 A cm^{-2} up to 10 T, well above H^* for films 2 and 3. Film 2 also exhibits substantial high-field current densities up to 20 K. J_c values of film 1 are well below those of film 3 at 4.2 and 20 K, which may be owing to its inhomogeneity. The bulk pinning-force, $F_p(H) = \mu_0 H J_c(H)$, exceeds 40 GN m^{-3} at 4.2 K for film 3, whereas F_p approaches 20 GN m^{-3} at 5 T for film 2 and remains above 10 GN m^{-3} between 1 and 10 T. These data are comparable to or exceed the pinning force of the established technological superconductors Nb47wt%Ti (ref. 6) and Nb₃Sn (ref. 7), which lie in the range $15\text{--}30 \text{ GN m}^{-3}$.

Figure 3c summarizes the irreversibility fields for bulk and thin-film MgB₂ derived from VSM data. The extrapolation to zero of the Kramer function ($J_c^{0.5} H^{0.25}$) was used to define the irreversibility field for the thin films. These extrapolations were nearly straight for films 1 and 3, just as in bulk samples², whereas the Kramer plots for film 2 were curved because of stronger pinning at higher fields. These analyses produce a $H^*(T)$ line for film 2 that has a much steeper slope, $\mu_0 dH^*/dT \approx 0.7 \text{ T K}^{-1}$, than the $H^*(T)$ data for bulk samples, which have a slope of about 0.3 T K^{-1} . This steeper $H^*(T)$ characteristic allows $H^*(T)$ for film 2 to intersect and exceed bulk sample $H_{c2}(T)$ values below about 10 K. In strong contrast to the film 2 data, the $H^*(T)$ analyses for lower resistivity films 1 and 3 lie along much lower lines, and do not exceed bulk $H^*(T)$ values at low temperature.

Transmission electron microscopy (TEM) was used to characterize the phase constitution and nanostructure of film 2. The MgB₂ grain size was remarkably small, approximately 10 nm, as shown in Fig. 4. Selected-area diffraction patterns were all ring patterns (Fig. 4 inset) and indicated the presence of significant volume fractions of both MgB₂ and MgO. As in the wavelength-dispersive microprobe results discussed earlier, energy-dispersive X-ray microanalysis in the TEM also suggests a considerably higher concentration of oxygen compared to what we have observed in TEM specimens from bulk MgB₂ (X.S. *et al.*, unpublished work).

The TEM analyses of film 2 and X-ray diffractometry of all three films indicate that all films have a [00l] texture with a large mosaic spread normal to the substrate and random in-plane orientation. The rocking curve widths of the (002) reflection of MgB₂ are 4.8° , 7.3° and 8.2° for films 1, 2 and 3, respectively. The broadening of the (002) X-ray peaks is consistent with the fine grain size observed by TEM. Using the Scherrer formula $d = 0.9 \lambda / B \cos \theta$, where B is the full width at half maximum of the 2θ diffraction peaks and λ is the X-ray wavelength used, grain dimensions along the c axis are 17 nm, 7 nm and 8 nm for films 1, 2 and 3, respectively. The MgO grains appeared

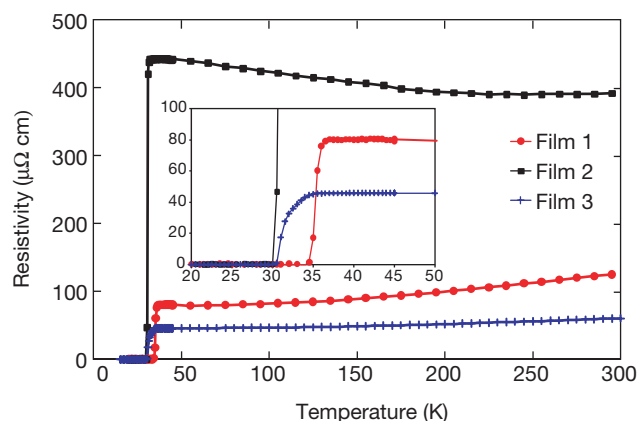


Figure 2 Resistivity as a function of temperature. Film 1 is indicated by the red curve with circles, film 2 by the black curve with squares, and film 3 by the blue curve with crosses. The inset shows a magnified view of the resistive transitions near the critical temperature. The uncertainty of the resistivity value is about $\pm 20\%$ owing to the large voltage contacts relative to sample dimensions.

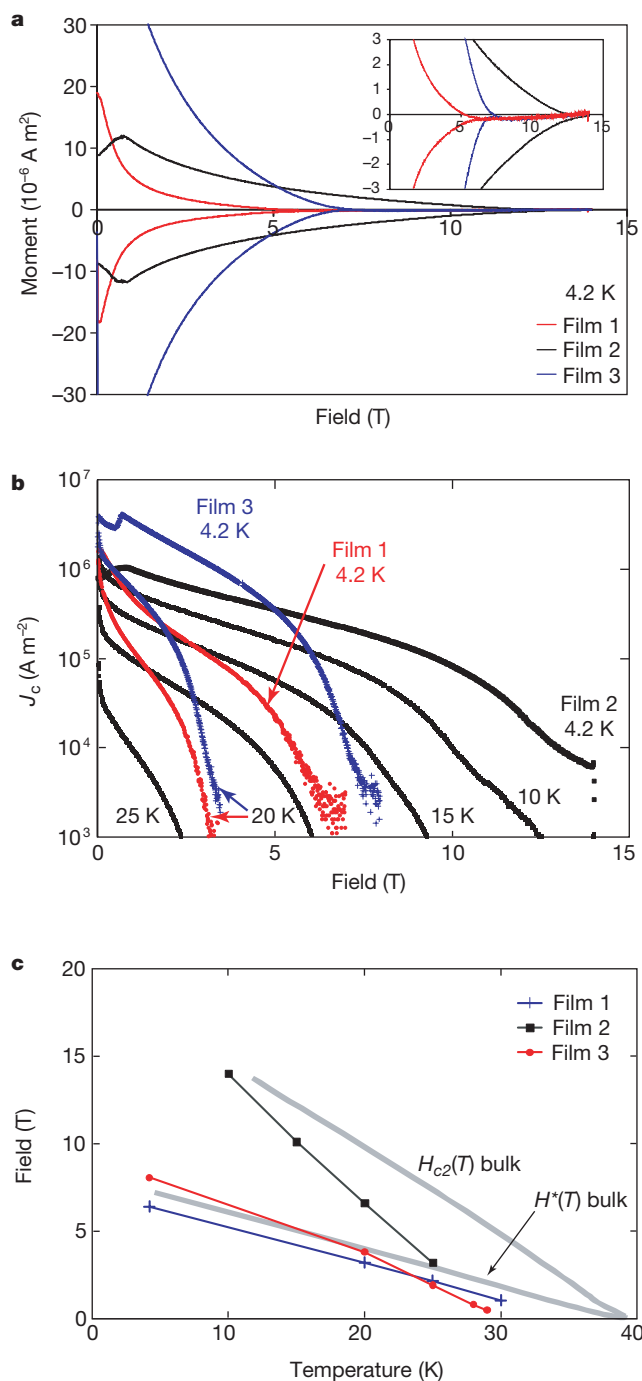


Figure 3 Magnetization measurements for films 1 (red curves), 2 (black curves), and 3 (blue curves). **a**, Raw magnetization data, obtained using a vibrating sample magnetometer, is compared at 4.2 K. The inset shows a magnified view of the closure of hysteresis loops, after subtracting a linear background that is due to the paramagnetic moment of the samples and the sample holder. **b**, The derived values of the critical current density are plotted as a function of field at 4.2 and 20 K for films 1 and 3, and at 4.2, 10, 15, 20, and 25 K for film 2. We note that the data for film 2 at 4.2 K are above 10^5 A cm^{-2} (a common benchmark for superconducting magnet applications) at nearly 10 T. **c**, The magnetization data are used to estimate an H - T phase diagram for MgB₂ thin films in comparison to bulk samples. The grey bands indicate established $H_{c2}(T)$ and $H^*(T)$ lines for bulk MgB₂, based on inductive^{2,4,5,9-11} and resistive^{4,5,9,11} measurements. The irreversibility fields for films 1 (red) and 3 (blue) lie close to the $H^*(T)$ line for bulk materials, even though their T_c values are lower than the approximate 39 K bulk value. The irreversibility field for film 2 increases at a faster rate of approximately 0.7 T K^{-1} below the critical temperature of 28.5 K, crossing the bulk $H_{c2}(T)$ line below about 10 K.

to be randomly oriented with a grain size of about 10 nm in all three films.

Taken as a whole, the experiment shows two primary results that have important implications for the potential of MgB_2 in superconducting magnets. First, unlike bulk samples, which appear to have a common T_c , $H_{c2}(T)$ and $H^*(T)$, here we see a strong enhancement of $H^*(T)$ that correlates well to the strongly enhanced resistivity and concomitant reduction of the electron mean free path and superconducting coherence length of film 2. Films 1 and 3, which were annealed in metal tubes and have lower oxygen concentrations, have metallic resistivity and smaller H^* values and $H^*(T)$ slopes. Film 2, which was annealed in quartz and has higher measured oxygen concentration and larger c axis parameter (0.3547 nm versus 0.3521 nm for bulk) has a weak increase of resistivity with decreasing temperature and a much steeper $H^*(T)$ curve, suggesting that oxygen can substitute for boron and expand the c axis of MgB_2 . In both alloy and intermetallic-compound superconductors, substitutional alloying to decrease the electron mean free path is very important for enhancing high-field superconductivity, as indicated by the direct dependence of $H_{c2}(0)$ on the normal-state resistivity through $\mu_0 H_{c2}(0) = 3.110 \rho \gamma T_c$ (ref. 14). Taking the coefficient of electronic specific heat $\gamma = 153 \text{ J m}^{-3} \text{ K}^{-2}$ for MgB_2 (ref. 15), the resistivity values predict $\mu_0 H_{c2}(0)$ values of 12, 57 and 7.6 T for films 1, 2 and 3 respectively, assuming that the dirty limit of the Ginsburg–Landau theory applies. Although we were not able to deduce H_{c2} values from the magnetization measurements, these predictions are at least qualitatively reasonable given the data in Fig. 3, which suggest extrapolated values of $\mu_0 H^*(0)$ of 10, 22 and 7 T for films 1, 2 and 3, respectively.

Second, the thin-film process produces textured MgB_2 films with extremely small grains and a substantial fraction of extremely small

(~ 10 nm) MgO particles. A high number density of flux pinning centres is crucial for obtaining useful ($> 10^5 \text{ A cm}^{-2}$) current densities at high fields. The exceptionally fine 10-nm scale of the grain size would be consistent with strong grain boundary pinning, as in Nb_3Sn , for which $F_p \propto d^{-1}$ (refs 16 and 17). Additional core-pinning by the non-superconducting MgO particles can also contribute to the overall pinning force¹⁷. Moreover, the very high J_c values confirm the absence of electromagnetic granularity². Evidently the threshold at which grain boundaries block current lies well above $J_c(T, H)$ and closer to the depairing current density of about 10^8 A cm^{-2} , an important requisite for attaining strong pinning in polycrystalline superconductors¹⁸.

In summary, it appears that MgB_2 can be alloyed, probably with oxygen, so as to increase the irreversibility field to levels where it may compete with today's materials of choice, $\text{Nb}_4\text{wt}\%\text{Ti}$ and Nb_3Sn . In addition to removing this fundamental limitation, the fine-grained films have technologically important current densities and strong flux pinning in perpendicular fields, which is the weaker orientation for layered superconductors. These results demonstrate that the principles that have guided the development of established high-field superconductors can also be applied successfully to MgB_2 . Improved high-field properties may thus be obtained in alloyed magnesium boride wires. We do not yet know how the development of texture, the extremely small grain size, the alloying of the MgB_2 , and the substantial presence of MgO are related, if at all. Further study of epitaxial MgB_2 and alloyed magnesium boride films on different substrate materials will be helpful in this regard. □

Received 20 March; accepted 27 April 2001.

- Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
- Larbalestier, D. C. et al. Strongly linked current flow in polycrystalline forms of the superconductor MgB_2 . *Nature* **410**, 186–189 (2001).
- Bud'ko, S. L. et al. Boron isotope effect in superconducting MgB_2 . *Phys. Rev. Lett.* **86**, 1877–1880 (2001).
- Finnemore, D. K., Ostenson, J. E., Bud'ko, S. L., Lapertot, G. & Canfield, P. C. Thermodynamic and transport properties of superconducting Mg^{10}B_2 . *Phys. Rev. Lett.* **86**, 2420–2422 (2001).
- Canfield, P. C. et al. Superconductivity in dense MgB_2 wires. *Phys. Rev. Lett.* **86**, 2423–2426 (2001).
- Lee, P. J. in *Wiley Encyclopedia of Electrical and Electronics Engineering* (ed. Webster, J. G.) Vol. 21, 75–87 (Wiley, New York, 1999).
- Smathers, D. B. in *Metals Handbook* 10th edn Vol. 2, *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials* 1060–1076 (ASM Intl, 1990).
- Slusky, J. S. et al. Loss of superconductivity with the addition of Al to MgB_2 and a structural transition in $\text{Mg}_{1-x}\text{Al}_x\text{B}_2$. *Nature* **410**, 343–345 (2001).
- Müller, K.-H. et al. The upper critical field in superconducting MgB_2 . Preprint cond-mat/0102517 at (xxx.lanl.gov) (2001).
- Takano, Y. et al. Superconducting properties of MgB_2 bulk materials prepared by high pressure sintering. *Appl. Phys. Lett.* **78**, 2914–2916 (2001).
- Bugoslavsky, Y., Perkins, G. K., Qi, X., Cohen, L. F. & Caplin, A. D. Critical currents and vortex dynamics in superconducting MgB_2 . *Nature* **410**, 563–565 (2001).
- Gyorgy, E. M., van Dover, R. B., Jackson, K. A., Schneemeyer, L. F. & Waszczak, J. V. Anisotropic critical currents in $\text{Ba}_2\text{YCu}_3\text{O}_7$ analyzed using an extended Bean model. *Appl. Phys. Lett.* **55**, 263–265 (1989).
- Kang, W. N., Kim, H.-J., Choi, E.-M., Jung, C. U. & Lee, S.-I. Epitaxial MgB_2 superconducting thin films with a transition temperature of 39 kelvin. *Science* (in press).
- Orlando, T. P., McNiff, E. J., Foner, S. & Beasley, M. R. Critical fields, Pauli paramagnetic limiting, and material parameters of Nb_3Sn and V_3Si . *Phys. Rev. B* **19**, 4545–4561 (1979).
- Wang, Y., Plackowski, T. & Junod, A. Specific heat in the superconducting and normal state (2–300 K, 0–16 T), and magnetic susceptibility of the 38-K superconductor MgB_2 : Evidence for a multi-component gap. Preprint cond-mat/0103181 at (xxx.lanl.gov) (2001).
- Dew-Hughes, D. The role of grain boundaries in determining J_c in high-field, high-current superconductors. *Phil. Mag.* **55**, 459–479 (1987).
- Campbell, A. M. & Evetts, J. E. Flux vortices and transport currents in type II superconductors. *Adv. Phys.* **21**, 199–428 (1972).
- Gurevich, A. & Cooley, L. D. Anisotropic flux pinning in a network of planar defects. *Phys. Rev. B* **50**, 13563–13576 (1994).

Acknowledgements

The work at the University of Wisconsin was supported by funding from the US Department of Energy, the Air Force Office of Scientific Research, the National Science Foundation through the Materials Research Science and Education Center for Nanostructured Materials and a David Lucile Packard Fellowship (C.B.E.). The work at Princeton University was supported by the National Science Foundation and the US Department of Energy.

Correspondence and requests for materials should be addressed to C.B.E. (e-mail: com@engr.wisc.edu).

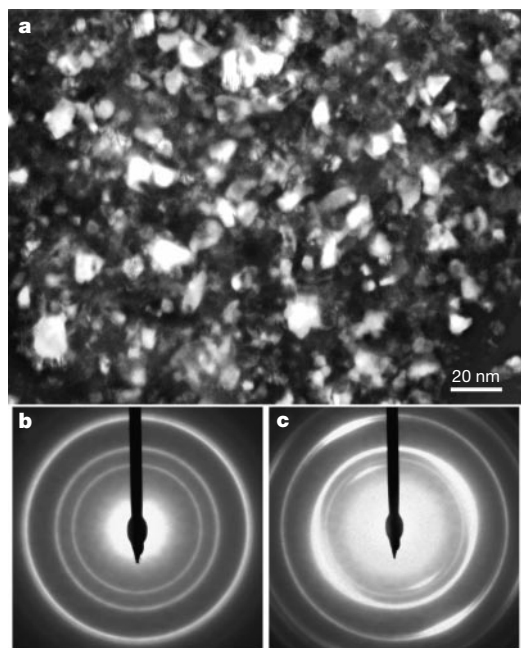


Figure 4 Transmission electron microscopy. **a**, Dark-field, plan-view image of film 2 in which only a fraction of the grains show bright intensity, clearly revealing the 10 nm grain size of the material. **b**, **c**, Selected area diffraction patterns from about $1 \mu\text{m}$ square areas of film 2. All of the rings can be attributed to either MgB_2 or MgO . The pattern in **b** was collected with the incident electron beam aligned approximately parallel to the film normal. It shows only complete rings and those rings that correspond to the MgB_2 phase are all of the $(hk0)$ type. The pattern in **c** was collected with the incident beam oriented at an angle to the film normal. It shows an uneven distribution of intensity along all MgB_2 rings. Considered together, these diffraction patterns indicate that the film possesses a $(00l)$ (that is, c axis) fibre texture oriented parallel to the film normal with little or no texture in the plane of the film.

Enhancement of the high-magnetic-field critical current density of superconducting MgB₂ by proton irradiation

Y. Bugoslavsky^{*†}, L. F. Cohen^{*}, G. K. Perkins^{*}, M. Polichetti^{*‡},
T. J. Tate^{*}, R. Gwilliam[§] & A. D. Caplin^{*}

^{*} Centre for High Temperature Superconductivity, Blackett Laboratory, Imperial College, London SW7 2BZ, UK

[†] General Physics Institute, 117942, Vavilov st38, Moscow, Russia

[‡] INFN-Dipartimento di Fisica, Università di Salerno, via S. Allende, Baronissi, I-84081 (Salerno), Italy

[§] EPSRC Ion Beam Centre, University of Surrey, Guildford, Surrey GU2 7RX, UK

Magnesium diboride, MgB₂, has a relatively high superconducting transition temperature¹, placing it between the families of low- and high-temperature (copper oxide based) superconductors. Supercurrent flow in MgB₂ is unhindered by grain boundaries^{2,3}, making it potentially attractive for technological applications in the temperature range 20–30 K. But in the bulk material, the critical current density (J_c) drops rapidly with increasing magnetic field strength⁴. The magnitude and field dependence of the critical current are related to the presence of structural defects that can 'pin' the quantized magnetic vortices that permeate the material, and a lack of natural defects in MgB₂ may be responsible for the rapid decline of J_c with increasing field strength³. Here we show that modest levels of atomic disorder induced by proton irradiation enhance the pinning of vortices, thereby significantly increasing J_c at high field strengths. We anticipate that either chemical doping or mechanical processing should generate similar levels of disorder, and so achieve

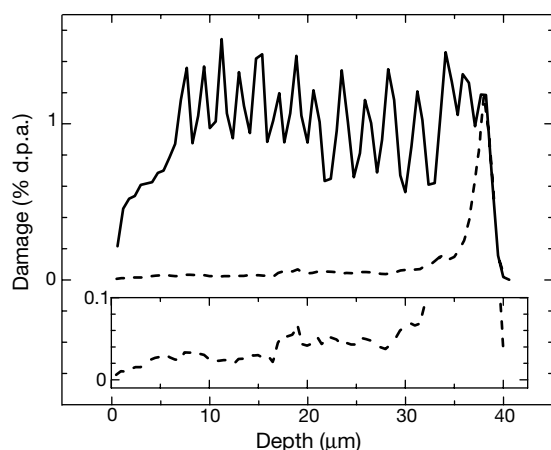


Figure 1 Simulated⁶ damage profiles produced in MgB₂ by proton irradiation. The damage is measured in units of displacements per atom (d.p.a.), but for protons at these energies the absolute values are known only to within a factor of about two. A monoenergetic proton beam causes an almost flat distribution of damage for most of its range, and then a sharp peak towards the end of the track. The broken line shows simulated damage for such a beam of energy 2 MeV and a fluence of 10^{16} cm^{-2} , as was used to irradiate the sample denoted 0.04% d.p.a. The inset shows the same data on an expanded scale. In order to create fairly uniform damage through the depth of a second sample, 15 consecutive implants were made with the beam energy varied between 2 MeV and 400 keV (solid line); the average damage is 1% d.p.a. A third sample was irradiated similarly, but a five times higher fluence, yielding 5% d.p.a. damage. In order to minimize sample heating, the beam current was maintained at $2 \mu\text{A cm}^{-2}$.

performance that is technologically attractive in an economically viable way.

A type II superconductor undergoes the transition to the normal state at the upper critical field, H_{c2} . However, the ability to carry dissipation-free current ceases at a lower field, the irreversibility field, H^* . Above H^* the Lorentz force on the vortices is large enough for them to become detached from pinning defects, and virtually free to move. In MgB₂ powder (and also wires and tapes), H^* is approximately half of H_{c2} (ref. 5), so that there is an extended field domain where there might be a useful J_c if only the pinning could be strengthened.

Ion irradiation is a reproducible means of inducing crystalline disorder. Energetic ions displace atoms from their equilibrium lattice sites, creating a variety of defects, including vacancies and interstitials. Such defects tend to depress the superconducting order parameter locally, and thereby create pinning sites for the vortices. We chose protons for this initial study because at the maximum beam energy available to us of 2 MeV, they can penetrate 40 to 50 μm into MgB₂. A series of irradiations were performed in order to create significant defect densities. The probability of displacement per atom (d.p.a.) can be simulated using commercial software⁶, and we aimed to generate as uniform a profile as possible through the sample depth (Fig. 1).

As in our previous study³, we selected fragments of about 100 μm size from commercial MgB₂ powder (Alfa Aesar Company, 98% purity). Several samples were prepared, each of 20 fragments embedded in silver-loaded epoxy and then polished, so that a defined geometry was obtained, with an estimated thickness for the MgB₂ fragments of about 50 μm . Also, this configuration ensured that good electrical and thermal contact with the conducting substrate was maintained during irradiation. Two samples were irradiated to fairly uniform damage levels of ~1% and 5% d.p.a. respectively, and one sample was more lightly damaged (~0.04% d.p.a. through most of its volume), but in a less uniform manner (see Fig. 1 legend). After the irradiation, the fragments were extracted from the epoxy and their magnetic moments were measured.

The transition temperature, T_c , drops with irradiation (Fig. 2). This could be caused by a number of possible factors: for example, depression of the density of states at the Fermi level by the disorder. Also, the transition broadens considerably in the irradiated samples; the inhomogeneity of the damage must contribute significantly to the breadth. It should be noted that in the 1 and 5% d.p.a. samples, the onset of the transition is lowered too, that is, there is no

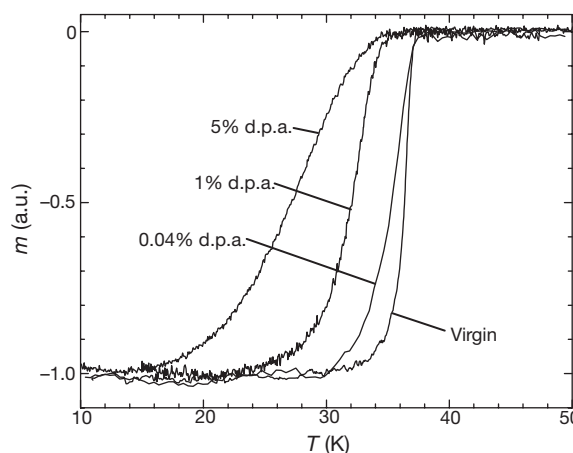


Figure 2 Effect of irradiation on the superconducting transition. As the damage increases, the onset transition temperature, T_c , decreases and the superconducting transition broadens, reflecting the disorder. The samples were zero-field cooled, a field of 10 mT applied, and the magnetic moment m measured while warming.

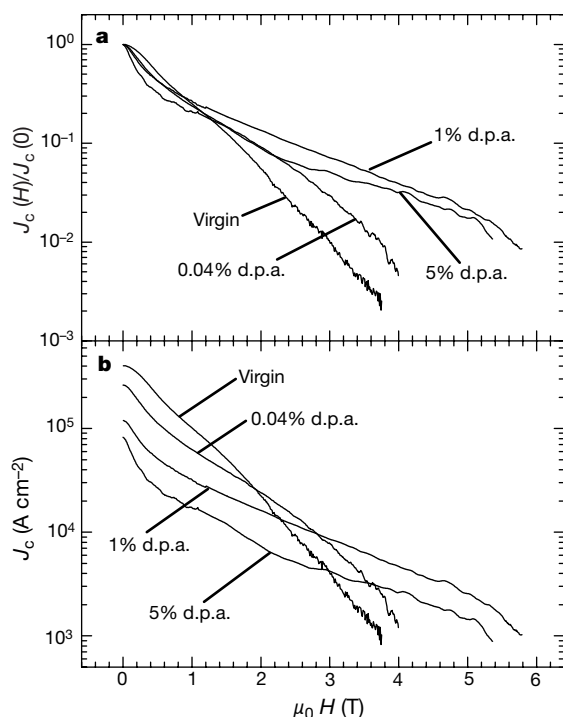


Figure 3 Effect of irradiation on the field-dependence of J_c at 20 K; the behaviour at other temperatures is similar. J_c is obtained from the magnetization hysteresis width, using the Bean model⁷. **a**, $J_c(H)$ of the virgin and irradiated samples at 20 K, with each sample normalized to its zero field value $J_c(0)$. Slower depression of J_c by the field, as occurs even with the lowest dose, signifies more efficient vortex pinning in the irradiated samples. At higher doses, the improvement in the gradient of $\log(J_c)-H$ curves saturates, but $J_c(0)$ is suppressed; therefore there is an optimal level of disorder for enhancement of J_c at high fields. **b**, Absolute values of J_c , which are somewhat uncertain, because the radiation damage is non-uniform. Also, in these polycrystalline fragments, it is possible that defects have migrated to grain boundaries, interrupting the circulation of supercurrent.

indication of the presence of any residual undamaged material in these samples.

We obtain the critical current density J_c from measurements of the irreversible magnetization in the standard fashion⁷. Irradiation does degrade J_c at low fields (Fig. 3), in part at least because of the reduction in T_c . However, the crucial result is that the field dependence of J_c flattens significantly in the irradiated samples. Even the lowest level of damage reduces the downward slope of the $\log(J_c)-H$ curve. Higher irradiation doses further decrease this slope by a factor of about 2. The combination of low-field degradation and high-field improvement yields a cross-over, so that, for example, the 1% d.p.a. sample at 20 K has a higher J_c than the virgin sample at magnetic fields above about 2.5 T.

Another way of quantifying the magnetic field dependence of J_c is by defining the irreversibility field, H^* , above which J_c drops below a useful level (Fig. 4). We show also the upper critical field, H_{c2} , for non-irradiated MgB_2 (ref. 5). Irradiation progressively increases the slope, dH^*/dT , and for the 5% d.p.a. sample it closely approaches the virgin sample dH_{c2}/dT . Very heavily damaged (by neutron irradiation) MgB_2 has been studied⁸: the disorder strongly reduces T_c , but leaves dH_{c2}/dT almost unchanged (J_c was not measured). It is therefore likely that dH_{c2}/dT is unchanged in our samples too.

Another report of an enhancement of H^* in MgB_2 is in thin (sub-micrometre) films deposited by laser ablation⁹. There too there is a depression of T_c , and the slope of dH^*/dT is even higher than in our irradiated samples (Fig. 4). These films are nanocrystalline, and it

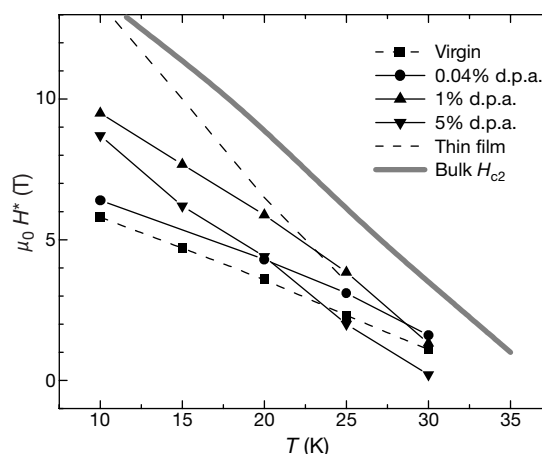


Figure 4 Temperature dependence of the irreversibility field H^* for different damage levels. The irreversibility field is that at which J_c becomes immeasurably small; here we use a criterion of 1 kA cm^{-2} . There is a systematic increase of the slope of $H^*(T)$ curve with increasing dose. For the 5% d.p.a. sample (down triangles), this slope closely approaches the slope of $H_{c2}(T)$ in the virgin sample (squares). The irreversibility field of the nanocrystalline MgB_2 thin film⁹ (dotted line) is shown too.

seems likely that in such films the grain boundaries are responsible for the strong vortex pinning.

At this stage, the detailed nature of the defects in irradiated MgB_2 is uncertain. The incident protons should be able to cause displacements on both Mg and B sublattices, and interstitials may form loops between the planes. At the end of their track, the protons may simply diffuse out, form MgH_2 , or react to form boranes. In any event, the defects form stronger pinning sites than those that are native in MgB_2 , as shown by the sharp reduction in slope of the J_c versus H plots.

Proton irradiation is clearly impractical for large-scale manufacture of MgB_2 high-current conductors, but these results demonstrate that modest disorder—at a level that should be achievable by chemical doping or alloying—in bulk MgB_2 can significantly improve J_c at high magnetic fields. The defect species and concentration will need to be optimized so as to minimize the reduction of T_c , while maximizing the enhancement of H^* .

The performance of high-temperature superconductors has improved steadily over 10 years of development⁴. The best commercially fabricated long-length tapes incorporating the $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_x$ phase achieve J_c values of about $1.5 \times 10^5 \text{ A cm}^{-2}$ at 20 K and zero applied field, dropping by a factor of about three in a field of 2 T. Both this absolute value of J_c and its (approximately exponential) field dependence are very close to those of irradiated MgB_2 (Fig. 3).

We suggest that at temperatures around 20 K, which can readily be achieved with standard cryocoolers, and at fields up to a few tesla, MgB_2 conductors with achievable pinning enhancement will be competitive in performance with high-temperature superconductors. The latter are costly to produce, so that if MgB_2 conductors can be fabricated cheaply—the constituents themselves are inexpensive—several applications of superconductivity, for example, for open-access medical magnetic resonance imaging magnets¹⁰, will become much more economically attractive. □

Received 2 April; accepted 30 April 2001.

1. Nagamatsu, J., Nakagawa, N., Muraoka, T., Zenitani, Y. & Akimitsu, J. Superconductivity at 39 K in MgB_2 . *Nature* **410**, 63–74 (2001).
2. Larbalestier, D. C. *et al.* Strongly linked current flow in polycrystalline forms of the new superconductor MgB_2 . *Nature* **410**, 186–189 (2001).
3. Bugoslavsky, Y., Perkins, G. K., Qi, X., Cohen, L. F. & Caplin, A. D. Vortex dynamics in superconducting MgB_2 and prospects for applications. *Nature* **410**, 563–565 (2001).
4. Vase, P., Flukiger, R., Leghissa, M. & Glowacki, B. A. Current status of high- T_c wire. *Supercond. Sci. Technol.* **13**, R71–R84 (2000).

5. Bud'ko, S. L. *et al.* Magnetoresistivity and complete $H_{c2}(T)$ in MgB_2 . *Phys. Rev. B* **63**, 220503(R) (2001).
6. Ziegler, J. F. *et al.* The Stopping and Range of Ions in Matter (SRIM-2000) (IBM Research, Yorktown NY, 2000).
7. Bean, C. P. Magnetization of high-field superconductors. *Rev. Mod. Phys.* **36**, 31–36 (1964).
8. Karkin, A. E. *et al.* Superconducting properties of MgB_2 probed by radiation-induced disordering. Preprint cond-mat/0103344 at (<http://xxx.lanl.gov>) (2001).
9. Eom, C. B. *et al.* High critical current density and enhanced irreversibility field in superconducting MgB_2 thin films. *Nature* **411**, 558–560 (2001).
10. Morrow, G. Progress in MRI magnets. *IEEE Trans. Appl. Supercond.* **10**, 744–751 (2000).

Acknowledgements

We are grateful to M. Stoneham and J. Matthews for advice on radiation damage. This work was supported by the UK Engineering and Physical Sciences Research Council.

Correspondence and requests for materials should be addressed to Y.B.
(e-mail: y.bugoslav@ic.ac.uk).

High critical currents in iron-clad superconducting MgB_2 wires

S. Jin, H. Mavoori, C. Bower & R. B. van Dover

Agere Systems/Lucent Technologies, Murray Hill, New Jersey 07974, USA

Technically useful bulk superconductors must have high transport critical current densities, J_c , at operating temperatures. They also require a normal metal cladding to provide parallel electrical conduction, thermal stabilization, and mechanical protection of the generally brittle superconductor cores. The recent discovery of superconductivity at 39 K in magnesium diboride (MgB_2)¹ presents a new possibility for significant bulk applications^{2–5}, but many critical issues relevant for practical wires remain unresolved. In particular, MgB_2 is mechanically hard and brittle and therefore not amenable to drawing into the desired fine-wire geometry. Even the synthesis of moderately dense, bulk MgB_2 attaining 39 K superconductivity is a challenge because of the volatility and reactivity of magnesium. Here we report the successful fabrication of dense, metal-clad superconducting MgB_2 wires, and demonstrate a transport J_c in excess of $85,000 A cm^{-2}$ at 4.2 K. Our iron-clad fabrication technique takes place at ambient pressure, yet produces dense MgB_2 with little loss of stoichiometry. While searching for a suitable cladding material, we found that other materials dramatically reduced the critical current, showing that although MgB_2 itself does not show the 'weak-link' effect characteristic of the high- T_c superconductors, contamination does result in weak-link-like behaviour.

We have made metal-clad wires by using the powder-in-tube technique that was used to make the Y–Ba–Cu–O oxide superconductor⁶. We note that a similar powder-in-tube approach has recently been used to fabricate metal-clad MgB_2 wires using various metals, such as Cu, Ag, Ni, Cu–Ni, or Nb (refs 7–9). In our study, the use of proper metal cladding has been found to be critical in producing well-sintered, high- J_c wires because of the strong chemical reactivity of MgB_2 . Magnesium in MgB_2 tends to react and combine with many metals such as Cu or Ag to form solid solutions or intermetallics with lowered melting points¹⁰ (as low as about 480 °C in the case of eutectic formation), which renders the metal cladding useless during sintering of MgB_2 at around 900–1,000 °C. Excluding elements with very low melting points such as Na or K, there are only a small number of essentially inert metals which exhibit no or little mutual solubility with Mg and do not form intermetallic compounds with Mg. These include Fe, Mo, Nb, V, Ta, Hf and W. Of these, the refractory metals (Mo, Nb, V, Ta, Hf, W), because of their greatly inferior ductility compared to Fe, are more

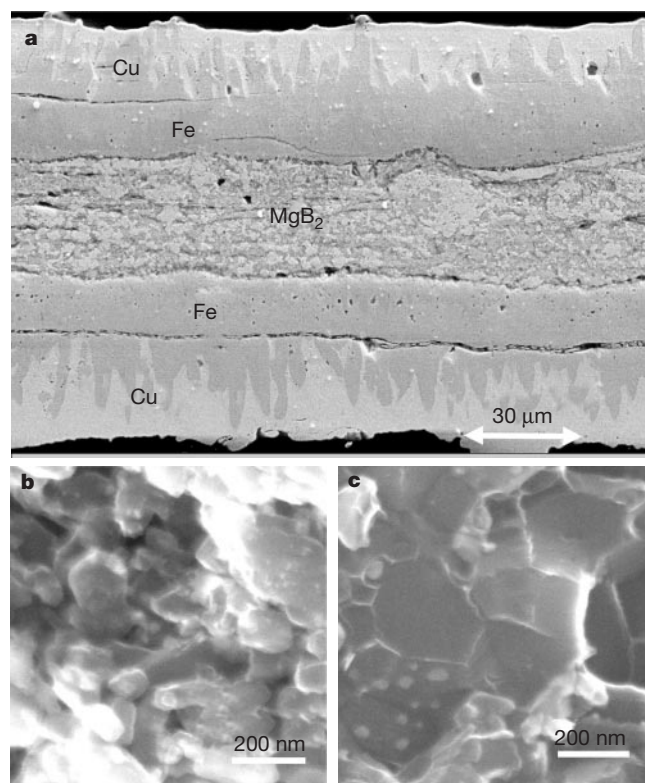


Figure 1 The MgB_2 ribbons were fabricated using Fe or Cu/Fe tubes with outside diameters of 5–6.35 mm. One end of the tube was sealed by crimping and the tube was then filled, in an argon atmosphere, with commercially available MgB_2 powder (98% purity, –325 mesh, from Alfa Aesar). The remaining end of the tube was also crimped and the composite structure was then swaged (and wire-drawn in some cases) followed by cold rolling to a ribbon geometry with 0.25–0.5 mm thickness, 3–5 mm width, and about 60 cm length. **a**, Longitudinal cross-sectional micrograph from the sintered ribbon. The superconductor core deforms continuously in conformation with the composite wire geometry during the deformation processes, presumably by particle–particle sliding. The Cu- and Fe-clad metal structure is well defined and distinguishable from the MgB_2 core which is around 35 μm thick. **b** and **c**, The high-magnification microstructures of the ribbons sintered (in argon) at 900 °C for 30 min and at 1,000 °C for 30 min, respectively. A dense structure with an ultrafine grain size of about 1,200 Å in average diameter is observed for the 900 °C sample. This is much smaller than the size of the starting MgB_2 powder material used (our SEM analysis gives an average of 3 μm), indicating substantial grain refinement by the wire fabrication process. Such grain refinement permits us to perform the necessary consolidation sintering at a lower temperature, thus reducing the extent of undesirable metallurgical reactions, such as the contamination of MgB_2 grain boundaries. The sample sintered at 1,000 °C gives a grain size about 2.5 times larger than that for the sample sintered at 900 °C. Unlike the Y–Ba–Cu–O type superconductors¹², the larger grain size in MgB_2 does not lead to a significant increase in critical currents, as indicated by our measurements of the magnetization J_c value.

difficult if not impossible to process with the substantial plastic deformation involved in fine-wire or ribbon fabrication. Thus Fe appears to be the best candidate material as a practical cladding metal or diffusion barrier for MgB_2 wire fabrication. Iron and its alloys are also inexpensive, naturally abundant and commercially widely available.

For MgB_2 wire/ribbon fabrication, we used mostly Fe tubes (or Cu tubes lined with an Fe inner tube as a diffusion barrier, Fig. 1a, which yielded essentially comparable superconducting characteristics). The tubes were filled with MgB_2 powder, drawn into wires, cold rolled to ribbons, and sintered. For direct and reliable measurements of T_c and J_c from the superconductor, the metal cladding was mechanically removed after wire fabrication

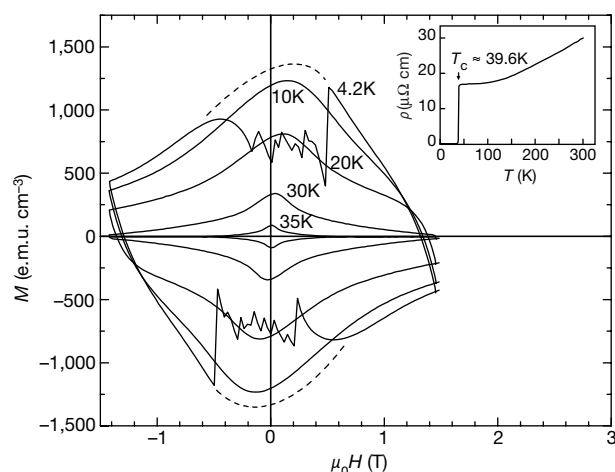


Figure 2 M - H curves for sintered iron-clad MgB_2 . The inset shows the resistivity versus temperature curve for MgB_2 from the iron-clad wire. The normal-state resistivity $\rho(40\text{ K})$ was about $17\ \mu\Omega\text{ cm}$ with a $\rho(40\text{ K})/\rho(298\text{ K})$ ratio of approximately $1/2$. These values are substantially different from those reported in ref. 3 on high-purity MgB_2 samples: $\rho(40\text{ K}) \approx 1\ \mu\Omega\text{ cm}$ and $\rho(40\text{ K})/\rho(298\text{ K}) \approx 1/20$. We note however, that the presence of some impurities is not necessarily undesirable for superconductors and is often beneficial for obtaining higher H_{c2} or enhanced J_c . The well known Bean model was used to yield the critical current as a function of field. For our square platelet samples, we used the formula $J_c = 30\Delta M/W$, where J_c is the critical current (in A cm^{-2}), ΔM is the difference between the upper and lower branches of the M - H curve (in e.m.u. cm^{-3}) and W is the transverse width of the sample (in cm). Data taken at 4.2 K show the effect of flux jumps (clear and sudden magnetization changes at fields below about $\mu_0 H < 0.5\text{ T}$), which occur only under conditions of extremely high critical currents and very low heat capacity and cause localized motion of magnetic flux¹³. This emphasizes the need for superconductor stabilization using normal metal cladding, as is standard for low- T_c superconductors. A crude interpolation of the data was employed to estimate J_c (magnetization) in the absence of the flux jumps (for example, when they were suppressed by normal metal cladding) as indicated by the dashed line. At 35 K , the M - H loop is completely closed for $\mu_0 H > 0.6\text{ T}$, that is, for H values greater than the irreversibility field at this temperature.

and high-temperature sintering. The bare MgB_2 ribbon so obtained was qualitatively dense, giving an audible 'ping' when dropped onto a hard surface or cut to smaller lengths using a hand-held wire cutter. The iron-clad bulk fabrication process we devised essentially maintains the intended stoichiometry of MgB_2 . For example, the weight loss during sintering of the ribbons at 900°C for 30 min in an argon atmosphere was measured to be only about 0.8% in terms of net MgB_2 weight change. By contrast, bare sintered pellets (prepared using a laboratory press at 3,000 bar compression, with the MgB_2 powder under argon atmosphere before and during the pressing), after the same 900°C sintering, lost around 31.4% of weight from MgB_2 (equivalent to an approximately 60% loss of magnesium). The sintered pellet was very porous and mechanically crumbly; it thus was not possible to make the electrical measurements, in agreement with the observation of ref. 11.

Figure 1a-c shows scanning electron microscopy (SEM) photographs illustrating the structure of the heat treated (sintered) composite, metal-clad wire or ribbon of MgB_2 . Substantial MgB_2 grain refinement seems to occur during the wire fabrication process by particle pulverization. It would be desirable to achieve higher levels of grain refinement (until the grain size is comparable to the MgB_2 coherence length of around $50\text{ \AA}$). For example, this might be done by optimized process control, by using finer starting powder, or by incorporating nanoscale, chemically inert dispersoid particles that would inhibit grain growth. Such grain refinement or presence

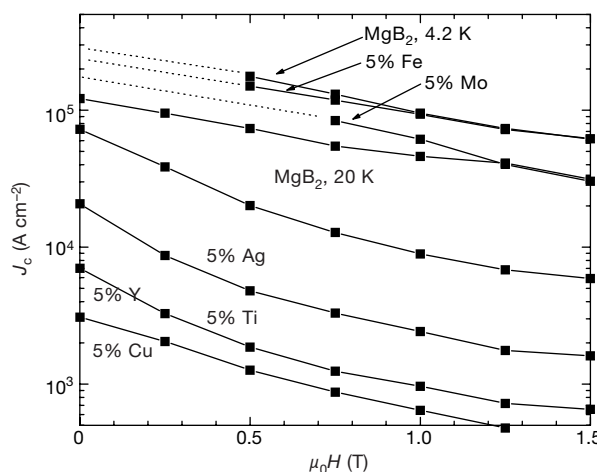


Figure 3 J_c (magnetization) versus H plots for various MgB_2 samples. The value for MgB_2 ribbon at 4.2 K and $H = 0$ is comparable to the J_c (transport) value from V - I curves on passing a pulse current of 30–100 A from a capacitor bank (about 1 ms rise time and about 8 ms decay time) and monitoring with a transient digitizer. The lead wires were attached using 10% In/Ag solder. A voltage criterion of $0.5\ \mu\text{V mm}^{-1}$ was used, primarily due to the limited voltage resolution in the digitizer. Because of the contact resistance, only a lower limit to the J_c (transport) could be measured. The J_c (magnetization) data for the various metal-doped MgB_2 were obtained at 4.2 K . We note that these metal additions strongly reduce the J_c (except iron). Increased amounts of metal additions (for example, 20 mol%), caused further reduction of J_c . It is possible that the atoms of the added metals, especially if they are molten (as is frequently observed for Cu or Ag cladding metal below their melting points during MgB_2 sintering), migrate by diffusion to the grain boundaries or to the MgB_2 particle surface before the completion of the sintering reaction. This may form a continuous or semi-continuous layer which is either nonsuperconducting or superconducting with reduced J_c . Cu, Ag and Y react with magnesium to form eutectic alloys while Fe, Mo, Y, Ag and Ti can react with boron to form intermetallics, presumably with different kinetics. The presence of such grain boundary layers would strongly impede the flow of supercurrents from grain to grain, causing the observed weak-link-like behaviour. Further research is needed to elucidate this aspect. Nonetheless, iron appears to be a particularly suitable material as one of the least detrimental cladding or diffusion-barrier metals for wire fabrication.

of nanodispersoids would be advantageous because it can facilitate introduction of defects and a higher density of flux-pinning centres. Also, it provides a useful way of introducing scattering centres so that the electron mean-free-path is reduced, thereby decreasing the coherence length and increasing the upper critical field.

The resistivity versus temperature curves for the bare MgB_2 samples stripped from the iron-clad ribbons were obtained by four-point measurement using a 10 mA a.c. current. As shown in the Fig. 2 inset, a sharp superconducting transition occurred at T_c (onset) $\approx 39.6\text{ K}$ with the T_c (midpoint) being around 38.4 K . Figure 2 shows typical M - H curves for MgB_2 (900°C samples), from which the J_c (magnetization) values are estimated. Figure 3 shows the J_c (magnetization) values so inferred as a function of magnetic field, for MgB_2 samples. The J_c (magnetization) values at 4.2 K were about $3 \times 10^5\text{ A cm}^{-2}$ at $H = 0$ and about $1 \times 10^5\text{ A cm}^{-2}$ at $H = 1\text{ T}$. At 20 K , the values were around $1.2 \times 10^5\text{ A cm}^{-2}$ at $H = 0$ and around $4 \times 10^4\text{ A cm}^{-2}$ at $H = 1\text{ T}$.

We have also evaluated transport J_c properties of our MgB_2 ribbon samples. In order to avoid the complications with the presence of normal metal cladding, the measurement was carried out with stripped (bare) MgB_2 ribbons with approximate dimensions of $0.1\text{--}0.3\text{ mm}$ thick $\times 1.2\text{ mm}$ wide $\times 13\text{ mm}$ long. At 4.2 K , the measurement gave a lower limit J_c (transport) $> 8.5 \times 10^4\text{ A cm}^{-2}$ for the MgB_2 ribbons. At 20 K the lower limit of J_c (transport) was measured to be

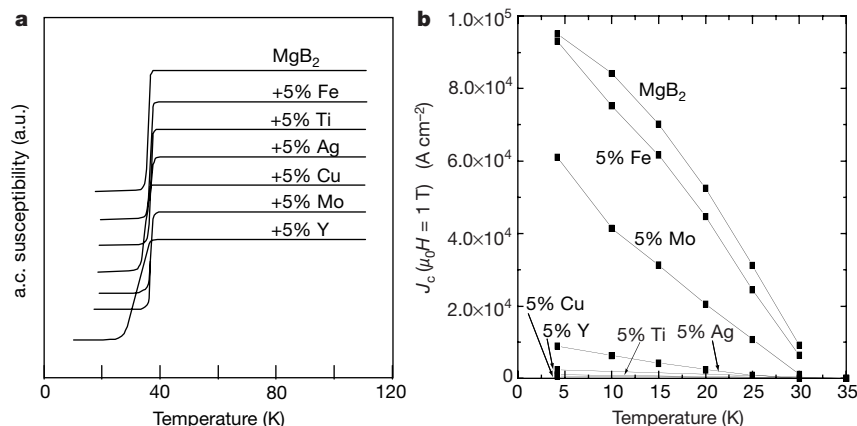


Figure 4 Effect of metal additions on the superconducting properties of MgB₂. **a**, Magnetic susceptibility versus temperature curves for MgB₂ containing fine metal particles. 5 mol% each of Fe, Mo, Cu, Ag, Ti (1–10 μ m average size), and Y (less than about 200 μ m) were thoroughly mixed (using mortar and pestle) with the MgB₂ powder, vacuum pumped, transferred into Fe tubes in argon, and the metal-clad ribbons fabricated and sintered (at 900 °C for 30 min) as described earlier. It should be mentioned that during the ribbon fabrication process used here, the thickness of these metal particles is also reduced substantially. Clearly the presence of these additions during sintering at 900 °C has little

effect on the T_c of MgB₂ (except for some broadening of the susceptibility transition in the case of Y, which has a considerable mutual solubility with Mg, forms a 566 °C eutectic, and reacts with B as well). The same trend was also confirmed with resistivity versus temperature measurements. This result suggests that the metal ions of Fe, Mo, Ag, Cu, Ti and Y are basically not incorporated into the MgB₂ superconductor phase. **b**, J_c (magnetization) versus temperature curves at a field of 1 T. The deterioration of J_c as a result of the metal additions occurs at all measurement temperatures.

2.3×10^4 A cm⁻². The measured transport J_c values are comparable to the zero-field J_c (magnetization) values, and will most probably approach the J_c (magnetization) values in future, better-controlled measurements with reduced contact resistance.

In view of the marginal upper critical field (H_{c2}) and flux-pinning characteristics in bulk MgB₂, it is important to improve such characteristics through structural or microstructural modifications for applications at 20 K or high temperatures. For example, chemical doping, introduction of precipitates or dispersoid particles, or atomic-scale control of defects such as vacancies, dislocations, grain boundaries, and so on may be attempted. In addition, possible effects of contamination of MgB₂ with other elements, such as from the cladding metal, is of interest and needs to be understood.

As part of such studies, we have evaluated the effect of several alloying metal elements on the critical current behaviour of the MgB₂ material. Although the T_c of MgB₂ remains unaffected by these elements (Fig. 4a), we notice significant alterations of critical current behaviour in these metal-containing MgB₂ samples as indicated in Fig. 3 (J_c versus H curves) and Fig. 4b (J_c versus temperature curves). The Fe addition appears to be least damaging, whereas the Cu addition causes J_c to be significantly reduced by 2–3 orders of magnitude with a slightly increased field dependence of J_c . These data clearly show that the inherently weak-link-free MgB₂ material is not necessarily exempt from ‘weak-link-like behaviour’, and any contamination of MgB₂ by foreign metal atoms on alloying or cladding needs to be carefully avoided or controlled. □

Received 27 March; accepted 3 May 2001.

1. Nagamatsu, J. *et al.* Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
2. Larbalestier, D. C. *et al.* Strongly linked current flow in polycrystalline forms of the superconductor MgB₂. *Nature* **410**, 186–189 (2001).
3. Finnemore, D. K. *et al.* Thermodynamic and transport properties of superconducting Mg¹⁰B₂. *Phys. Rev. Lett.* **86**, 2420–2422 (2001).
4. Canfield, P. C. *et al.* Superconductivity in dense MgB₂ wire. *Phys. Rev. Lett.* **86**, 2423–2426 (2001).
5. Eom, C. B. *et al.* High critical current density and enhanced irreversibility field in superconducting MgB₂ thin films. *Nature* **411**, 558–560 (2001).
6. Jin, S. *et al.* High T_c superconductors—composite wire fabrication. *Appl. Phys. Lett.* **51**, 203 (1987).
7. Sumption, M. D. *et al.* Transport current in MgB₂-based superconducting strand at 4.2 K and self-field. Preprint cond-mat/0102441 at (xxx.lanl.gov) (2001).
8. Grasso, G. *et al.* Large transport critical currents in unsintered MgB₂ superconducting tapes. Preprint cond-mat/0103563 at (xxx.lanl.gov) (2001).

9. Glowacki, B. A. *et al.* Superconductivity of powder-in-tube MgB₂ wire. *Supercond. Sci. Technol.* **14**, 193–199 (2001).
10. Masalski, T. (ed.) *Binary Phase Diagrams*. 46, 1,434 2nd edn (ASM International, Metals Park, OH, 1990).
11. Slusky, J. S. *et al.* Loss of superconductivity with the addition of Al to MgB₂ and a structural transition in Mg_{1-x}Al_xB₂. *Nature* **410**, 343–345 (2001).
12. Jin, S. *et al.* High critical currents in Y-Ba-Cu-O superconductors. *Appl. Phys. Lett.* **52**, 2074–2076 (1988).
13. Tinkham, M. *Introduction to Superconductivity* 185 (McGraw-Hill, New York, 1975).

Correspondence and requests for materials should be addressed to S.J. (e-mail: jin@agere.com).

Hyperconjugation not steric repulsion leads to the staggered structure of ethane

Vojislava Pophristic & Lionel Goodman

Wright and Rieman Chemistry Laboratories, Rutgers University, New Brunswick, New Jersey 08903, USA

Many molecules can rotate internally around one or more of their bonds so that during a full 360° rotation, they will change between unstable and relatively stable conformations. Ethane is the textbook example of a molecule exhibiting such behaviour: as one of its two methyl (CH₃) groups rotates once around the central carbon–carbon bond, the molecule will alternate three times between an unstable eclipsed conformation and the preferred staggered conformation. This structural preference is usually attributed to steric effects^{1–7}; that is, while ethane rotates towards an eclipsed structure, the electrons in C–H bonds on the different C atoms are drawing closer to each other and therefore experience increased repulsion, introducing a rotation barrier that destabilizes the eclipsed structure^{8,9}. Stabilization of the staggered structure through rotation-induced weakening of the central C–C bond¹⁰ and hyperconjugation^{11,12} has been considered to be

involved, but evaluation of the contributions of these effects to ethane's internal rotation barrier and conformational preference remains difficult^{13,14}. Here we report a series of ethane structure optimizations, where successive removal of different interactions indicates that ethane's staggered conformation is the result of preferential stabilization through hyperconjugation. Removal of hyperconjugation interactions yields the eclipsed structure as the preferred conformation, whereas repulsive forces, either present or absent, have no influence on the preference for a staggered conformation.

Here we consider the role of three principal physical factors—exchange, electrostatic and hyperconjugative interactions—underlying ethane's structural preference. Both hyperconjugation and exchange repulsion are quantum mechanical effects arising from orbital overlap. The former involves electron transfer from an occupied to an unoccupied orbital, leading to delocalization of charge (Fig. 1b). The latter involves the Pauli exclusion principle, which requires pairs of electrons not to occupy the same spatial region. In contrast to these two interactions, electrostatic or Coulomb interaction involves classical $1/R$ repulsion between charges. Exchange repulsion, because of the orbital overlap requirement, is necessarily a short-range effect, whereas there is no such requirement for the Coulomb interaction.

First, we considered whether Pauli exchange and electrostatic (coulombic) repulsions are factors leading to the staggered structure. Taken together, these are frequently regarded as the steric factor. To come to a decisive conclusion concerning the role of exchange repulsion, two cases are considered in Fig. 2: the real ethane molecule with all interactions present, and a hypothetical ethane with exchange repulsion removed. We emphasize that total exchange repulsion is considered because it is well established that limited repulsion between pairs provides a poor approximation to total exchange repulsion, largely because of a domino effect on the other electrons¹⁵. Thus, both exchange and Coulomb repulsions must be considered as all-electron phenomena. Potential curves

were constructed in terms of the torsional angle, one curve (with all interactions present) representing the energy of the real system, the other, the energy of the system with exchange repulsion absent (Fig. 2). The conclusion from Fig. 2 is that the minima of both curves coincide, establishing that the staggered conformation is preferred regardless of the presence of exchange repulsion.

The effect of Coulomb repulsion on the orientation of the methyl groups can be estimated by comparing its torsional angle dependence for three models: rigid rotation (that is, pure rotation, all skeletal flexings frozen), partially relaxed rotation (C–C bond lengthening included) and fully relaxed rotation (all skeletal flexings taken into account). For the rigid rotation model, both electron and nuclear Coulomb repulsions increase as the torsional angle incrementally increases to the eclipsed conformer (Fig. 3). When C–C bond lengthening is included as part of the rotational process, however, both repulsions decrease. Thus, central bond stretching reduces the strain that is accumulated in the molecule by rotation alone, leading to a large decrease in electrostatic repulsion energies. Coulomb repulsion therefore does not explain the staggered structural preference in this molecule.

We next approach the hyperconjugative interactions (inherently stabilizing; Fig. 1b) by examining the internal rotation-induced change in nuclear–electron attraction. It is this component of the potential energy, ΔV_{ne} (as opposed to repulsive nuclear–

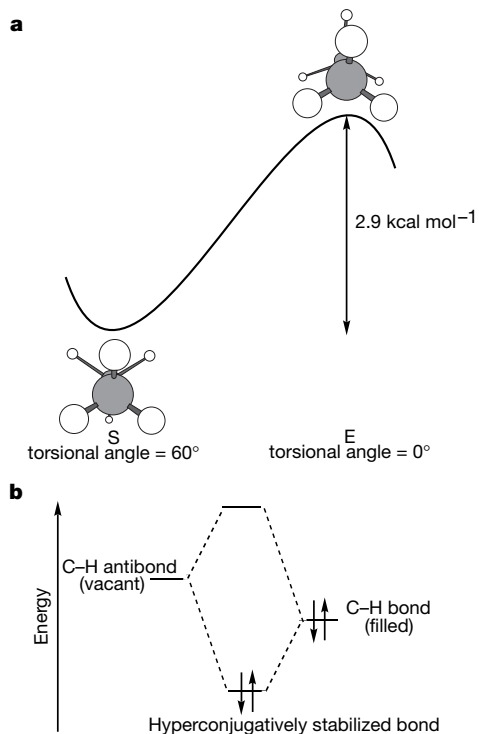


Figure 1 Ethane energetics. **a**, Ethane staggered (S) and eclipsed (E) conformers. **b**, Vicinal hyperconjugative stabilization by overlap between an occupied and an unoccupied orbital on two methyl groups.

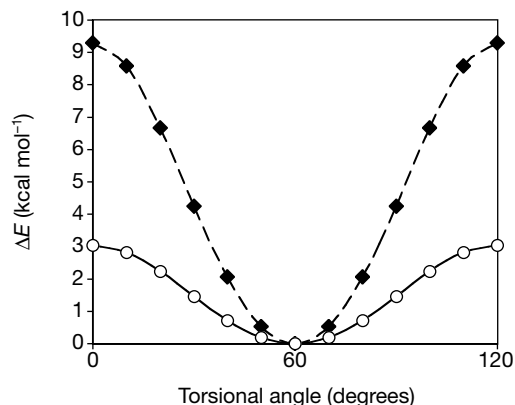


Figure 2 Torsional angle dependencies of energy of real ethane (solid line) and a hypothetical ethane with exchange repulsion absent (dashed line).

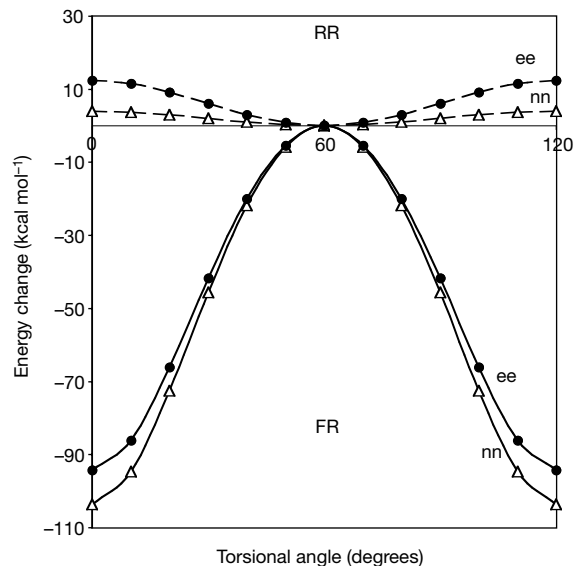


Figure 3 Torsional angle dependence of nuclear–nuclear (nn) and electron–electron (ee) repulsion energy changes for rigid (RR) and fully relaxed (FR) rotation.

Table 1 Partitioning of nuclear–electron potential (ΔV_{ne}) for various skeletal relaxations

$D_{3d}(S) \rightarrow D_{3h}(E)$	Step I	Step II	Step III	FR
<i>a</i> category (σ)	4.1	153.1	–1.9	154.8
<i>e</i> category (π)	–26.1	69.8	4.3	49.0

Step I, rigid rotation (molecular skeleton frozen throughout rotation); step II, C–C bond lengthening alone (no rotation); step III, rotation including CH_3 group flexing alone; FR, fully relaxed rotation, including all skeletal relaxations. HF/6-311G(3df,2p) energy (kcal mol^{–1}), using MP2/6-311G(3df,2p) geometry.

nuclear and electron–electron interactions) that connects most directly to the electron transfers inherent in hyperconjugative interactions. Insight into those factors that control the increase can be obtained by dissecting ΔV_{ne} into its symmetry components, which ultimately can be categorized as σ and π contributions. The 18-electron molecular orbital configuration appropriate to the staggered conformation D_{3d} symmetry is $(1a_{1g})^2(1a_{2u})^2(2a_{1g})^2(2a_{2u})^2(e_g)^4(3a_{2u})^2(e_u)^4$ and for the D_{3h} symmetry eclipsed conformer, $(1a_1')^2(1a_2'')^2(2a_1')^2(2a_2'')^2(e'')^4(3a_2'')^2(e')^4$. In this respect, ΔV_{ne} contributions from *a* category orbitals represent the σ interactions and those from *e* category orbitals represent π interactions. The higher-lying a_{1g} and a_{2u} orbitals have large carbon p_σ and valence shell s contributions and contribute importantly to the C–C bond strength. The two degenerate orbitals are very different from the σ ones in their composition; they have large hydrogen atom s character. These orbitals most importantly connect to the hyperconjugative delocalization interactions between the methyl groups. As can be seen from Table 1, the overall effect on ΔV_{ne} by fully relaxed rotation to the eclipsed conformer is a large increase, that is, the attraction is decreased—this conclusion holds for both $\Delta V_{ne}(\sigma)$ and $\Delta V_{ne}(\pi)$.

Table 1 also separates the skeletal and methyl-folding relaxations from the torsional rotation of the methyl group¹⁶. For step I (rigid rotation step), the σ symmetry classes (these involve C–C bond strength) contribute little to ΔV_{ne} ; for step II, the equilibrium conformer expansion step, both σ and π contributions to ΔV_{ne} are large (but σ dominates). Step III, which involves methyl-group folding alone, is negligible. There are some clear conclusions. $\Delta V_{ne}(\sigma)$ for the equilibrium conformer expansion step (II) (with no rotation) mimics $\Delta V_{ne}(\sigma)$ for fully relaxed rotation. In effect, the pure rotation model places undue emphasis on the hyperconjugative interactions by neglecting weakening of the C–C bond that controls the σ orbital contribution to ΔV_{ne} . On the other hand, for the rigid rotation step (step I) the π classes (*e* categories) are strongly active. In ethane, there are only two types of hyperconjugative interactions: vicinal (between methyl groups) and geminal (within a single methyl group). Only π orbitals have the overlap between the two methyl groups necessary for vicinal hyperconjugation. The overall conclusion is that a full molecular orbital analysis of the ΔV_{ne} term shows the importance of hyperconjugative interactions to the rotational barrier, independent of any localized orbital model.

Natural bond orbitals (NBOs)¹⁷ provide the means to analyse the hyperconjugative interactions in more detail than is possible by molecular orbitals. These orbitals are nearly doubly occupied, localized bonding orbitals (bonds) and nearly vacant, localized antibonding orbitals (antibonds). An individual hyperconjugative interaction is expressed as a charge (electron) transfer between

selected bonds and antibonds, allowing specific hyperconjugative interactions that influence structural preference to be pinpointed. Following this paradigm, the ethane structure is examined through geometry optimizations with selected charge transfers absent (Table 2). Thus, if inversion of conformational preference occurs on removal of a specific charge transfer interaction, the controlling factor for ethane's staggered structure has been identified. In our reasoning, natural bond orbitals are not used to provide the evidence for hyperconjugative preferential stabilization of the staggered conformer. Rather, their value lies in providing insight into which charge transfer interactions are the controlling ones by partitioning the full hyperconjugative interaction into its component parts, vicinal and geminal. The 'no hyperconjugation' entry in Table 2 shows that preference for the staggered conformation is lost on removal of all charge transfers. Furthermore, the equilibrium conformer of ethane with only the vicinal hyperconjugation removed ('no vicinal hyperconjugation' entry) is eclipsed. If only geminal hyperconjugation is deleted, inversion of conformational preference does not occur. The conclusion is that vicinal charge transfer interactions are the ones that keep the molecule in the staggered conformation. By this argument, it is the hyperconjugatively induced preferential stabilization of the staggered conformer that controls the ethane structure.

The important conclusion, obtained from our hyperconjugation and repulsion deletion calculations, is that ethane adopts the eclipsed conformation when the vicinal charge transfers are absent, and that no inversion of structure occurs on removal of coulombic and exchange repulsions. This, along with symmetry analysis of the full molecular orbital wavefunction, shows decisively that it is hyperconjugation, not repulsive forces, that determines the equilibrium structure of ethane.

These direct structural optimizations prove that vicinal hyperconjugation determines the staggered structure and validates Brunck and Weinhold's structural inference derived from barrier analysis in 1979 (ref. 11). A central point in eliminating repulsive interactions as a factor in determining the staggered structure is the inclusion of skeletal expansion, explicitly C–C bond lengthening—that is, internal rotation is not pure rotation. The conclusions introduced here are expected to have an important role in analysing such diverse molecular phenomena as polymer rheology and protein folding. □

Methods

Natural bond orbitals

NBOs used here are two-centred orbital transforms of HF/6-311G(3df,2p) and HF/6-311G(3df,3pd) molecular orbitals into almost doubly occupied bonding orbitals (bonds) and nearly unoccupied antibonding orbitals (antibonds)¹⁷. Unlike virtual orbitals in molecular orbital theory, which are entirely vacant and thus have no involvement in observable properties, antibonds have a low residual occupancy and participate in energy effects. Owing to their localized nature, the NBO representation of ethane corresponds closely to the Lewis structure. The molecular orbital notation defines the basis set and method used in the calculation. For example, the 6-311G(3df,2p) represents a linear combination of gaussian functions utilized to form the orbitals. HF refers to the Hartree–Fock self-consistent field computational method.

Exchange repulsion

Exchange repulsion arises as a consequence of the Pauli-principle requirement that the N-electron wavefunction be antisymmetrical with respect to interchange of pairs of electrons. In effect, wavefunction antisymmetry provides the 'quantum pressure' that resists crowding too many electrons into the same spatial region¹⁸. The principal energetic consequence of antisymmetrization is the (implicit) orbital orthogonalization¹⁹. The curve that excludes exchange repulsion in Fig. 2 is calculated from the energetic cost of orthogonalization of non-orthogonal NBOs (HF/6-311G(3df,2p) energy, using MP2/6-311G(3df,2p) geometries).

Hyperconjugation and energy partitioning

In the NBO description, hyperconjugation represents electron transfer between bonds and antibonds. The effect of hyperconjugation interactions was obtained by deletion of selected antibonds in the NBO description of ethane. We obtained the geometry optimizations with these deletions using the very tight optimization option of the NBO 4.0 (ref. 20) module of the Gaussian 98 (ref. 21) software package.

Table 2 Conformational dependence on hyperconjugative interactions

Deleted hyperconjugative interactions	Torsional angle (degrees)	Corresponding conformer
No deletion	60.0	S
No hyperconjugation	0	E
No vicinal hyperconjugation	0	E
No geminal hyperconjugation	60.0	S

S, staggered conformer; E, eclipsed conformer. HF/6-311G(3df,3pd) geometry optimization. These trends are found to be independent of basis set.

The total electronic energy change (ΔE) for any of the processes defined in Table 1 contains kinetic (ΔT), electron–electron (ΔV_{ee}), nuclear–nuclear (ΔV_{nn}) and nuclear–electron (ΔV_{ne}) potential energy terms: $\Delta E = \Delta T + \Delta V_{ee} + \Delta V_{nn} + \Delta V_{ne}$. Only the attractive nuclear–electron potential (ΔV_{ne}) is involved in hyperconjugative interactions. These have been broken down according to molecular orbital symmetry classes. Coulomb repulsions are contained in both ΔV_{ee} and ΔV_{nn} .

Received 22 September 2000; accepted 29 March 2001.

1. Jones, M. Jr *Organic Chemistry* (W. W. Norton & Company, New York, 2000).
2. Carey, F. A. *Organic Chemistry* (McGraw Hill, New York, 2000).
3. McMurry, J. *Organic Chemistry* (Brooks/Cole, Thomson Learning, New York, 1999).
4. Vollhardt, K. P. C. & Schore, N. E. *Organic Chemistry: Structure and Function* (W. H. Freeman and Company, New York, 1998).
5. March, J. *Advanced Organic Chemistry* (John Wiley & Sons, New York, 1992).
6. Loudon, G. M. *Organic Chemistry* (Benjamin/Cummings, Menlo Park, California, 1988).
7. Carey, F. A. & Sundberg, R. J. *Advanced Organic Chemistry* (Kluwer Academic/Plenum, New York, 2000).
8. Sovers, O. J., Kern, C. W., Pitzer, R. M. & Karplus, M. Bond-function analysis of rotational barriers: ethane. *J. Chem. Phys.* **49**, 2592–2599 (1968).
9. Christiansen, P. A. & Palke, W. E. A study of the ethane internal rotation barrier. *Chem. Phys. Lett.* **31**, 462–466 (1975).
10. Bader, R. F. W., Cheeseman, J. R., Laidig, K. E., Wiberg, K. B. & Breneman, C. Origin of rotation and inversion barriers. *J. Am. Chem. Soc.* **112**, 6530–6536 (1990).
11. Brunk, T. K. & Weinhold, F. Quantum mechanical studies on the origin of barriers to internal rotation about single bonds. *J. Am. Chem. Soc.* **101**, 1700–1709 (1979).
12. Reed, A. E. & Weinhold, F. Natural bond orbital analysis of internal rotation barriers and related phenomena. *Isr. J. Chem.* **31**, 277–285 (1991).
13. Wiberg, K. B. & Rablen, P. R. Comparison of atomic charges derived via different procedures. *J. Comp. Chem.* **14**, 1504–1518 (1993).
14. Reed, A. E. & Weinhold, F. Some remarks on the C–H bond dipole moment. *J. Chem. Phys.* **84**, 2428–2430 (1986).
15. Badenhoop, J. K. & Weinhold, F. Natural steric analysis of internal rotation barriers. *Int. J. Quant. Chem.* **72**, 269–280 (1999).
16. Goodman, L., Gu, H. & Pophristic, V. Flexing analysis of ethane internal rotation energetics. *J. Chem. Phys.* **110**, 4268–4275 (1999).
17. Weinhold, F. in *The Encyclopedia of Computational Chemistry* (ed. Schleyer, P. v. R.) 1792–1811 (John Wiley & Sons, Chichester, 1998).
18. Weisskopf, V. F. Of atoms, mountains, and stars: a study in qualitative physics. *Science* **187**, 605–612 (1975).
19. Badenhoop, J. K. & Weinhold, F. Natural bond orbital analysis of steric interactions. *J. Chem. Phys.* **107**, 5406–5421 (1997).
20. Glendening, E. D., Badenhoop, J. K., Reed, A. E., Carpenter, J. E. & Weinhold, F. NBO 4.0 (Theoretical Chemistry Institute, University of Wisconsin, Madison, 1996).
21. Frisch, M. J. *et al.* Gaussian 98 (Gaussian, Pittsburgh, Pennsylvania, 1998).

Acknowledgements

We thank R. Sauer for comments. Support by the National Science Foundation is acknowledged.

Correspondence and requests for materials should be addressed to L.G. (e-mail: goodman@rutchem.rutgers.edu).

Possible displacement of the climate signal in ancient ice by premelting and anomalous diffusion

A. W. Rempel*, E. D. Waddington†, J. S. Wettlaufer*‡ & M. G. Worster§

* Applied Physics Laboratory, University of Washington, Box 355640, Seattle, Washington 98105, USA

† Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA

‡ Department of Physics, University of Washington, Box 351560, Seattle, Washington 98105, USA

§ Institute of Theoretical Geophysics, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK

The best high-resolution records of climate over the past few hundred millennia are derived from ice cores retrieved from Greenland and Antarctica^{1–3}. The interpretation of these records relies on the assumption that the trace constituents used as

proxies for past climate have undergone only modest post-depositional migration. Many of the constituents are soluble impurities found principally in unfrozen liquid that separates the grain boundaries in ice sheets. This phase behaviour, termed premelting, is characteristic of polycrystalline material^{4,5}. Here we show that premelting influences compositional diffusion in a manner that causes the advection of impurity anomalies towards warmer regions while maintaining their spatial integrity. Notwithstanding chemical reactions that might fix certain species against this prevailing transport, we find that—under conditions that resemble those encountered in the Eemian interglacial ice of central Greenland (from about 125,000 to 115,000 years ago)—impurity fluctuations may be separated from ice of the same age by as much as 50 cm. This distance is comparable to the ice thickness of the contested sudden cooling events in Eemian ice from the GRIP core.

At the top of an ice sheet, in the firn layer (typically a few tens of metres thick⁶), some trace constituents are displaced by vapour transport through the connected network of air pockets that

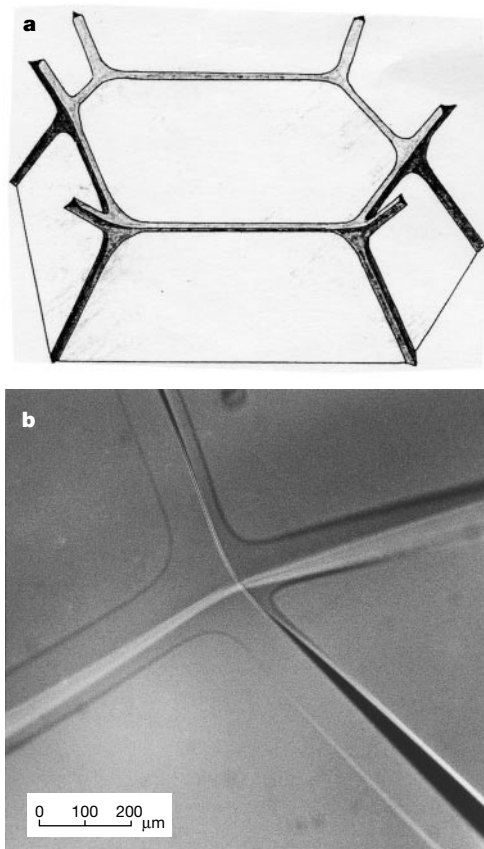


Figure 1 The hydraulic system of polycrystalline ice. **a**, A three-dimensional schematic diagram of the vein–node system (taken from ref. 10, where it was modified from ref. 25). For the dihedral angles observed in ice, the liquid is connected in a continuous network⁹. **b**, A photograph (H. M. Mader, personal communication; see ref. 11) of four veins intersecting at a node between four ice grains near 0 °C. Several mechanisms are responsible for the existence of liquid at temperatures in the solid region of the bulk phase diagram—known as premelting^{4,5}—these include the effects of impurities and interfacial curvature, which cause liquid to line the boundaries where three or more grains meet. (Premelted films can form on the surfaces of dust particles, bubbles and clathrate crystals; for typical volume fractions of these constituents, the mass fraction of liquid sequestered in films on their surfaces is expected to be much smaller than the mass fraction of liquid in the vein–node system (see Supplementary Information). Inter-molecular interactions between adjacent pairs of grains^{4,5} have been invoked to explain laboratory observations of grain-boundary melting²⁶; the prevalence of this behaviour in the polar ice sheets has not been measured.)

separate individual ice grains, and seasonal melting can sometimes cause additional advective transport⁷. Beneath the firn, the air pockets are sealed and, at the cold temperatures that prevail within the polar ice sheets, the flow of liquid is likely to be negligible. As solid-state diffusion through single ice grains is extremely slow⁸, the impurities are normally considered to be 'frozen' in place. However, the ubiquitous presence of liquid water along the boundaries where three ice grains meet⁹ provides an alternative route for diffusive transport. The veins, which separate three adjacent grains, and nodes, which connect the veins at junctions between four grains, form a continuous network of microscopic channels that remain liquid at sub-zero temperatures^{9–11} (Fig. 1). Diffusion through the unfrozen liquid within an ice sheet can enhance the exchange of oxygen isotopes¹², present in both the solid and liquid phases, and explains why the amplitude of seasonal variations in the isotopic ratio is reduced much more rapidly than solid-state diffusion can accomplish¹³. Here we model the transport of soluble impurities, which control the fraction of premelted liquid, and explore how records of the bulk concentrations of these climate indicators may have been altered over millennial timescales.

Most impurity molecules are too large to be incorporated within the ice lattice⁷. Instead they are expected to be confined to grain boundaries, either as solid precipitates or, if the temperature is above the eutectic, as dissolved constituents of the liquid phase. At

present, the average temperature is roughly -32°C at the top of the GRIP and GISP2 ice cores in central Greenland^{14,15} and -55°C at Vostok in Antarctica². Even these frigid temperatures are much higher than the eutectic temperatures of some of the aqueous solutions that are likely to be present. For instance, aqueous solutions of sulphuric acid (H_2SO_4), associated with the fallout from volcanic eruptions, have a eutectic temperature of about -73°C and the eutectic temperature of ammonium hydroxide, tied to periods during which vast continental regions were engulfed in massive forest fires, is roughly -84°C (ref. 16). Direct evidence for the presence of concentrated acid solutions along the veins in Antarctic ice samples was provided by measurements of sulphur concentrations¹⁷, thought to have derived from H_2SO_4 and the combination of sulphate and nitrate concentrations¹⁸, thought to represent quantities of H_2SO_4 and nitric acid (HNO_3). In order to display the fundamental effect of anomalous diffusion we treat the idealized case in which the liquid contains a single species of dissolved impurities. Because the most prevalent impurity species are often observed to co-vary through large sections of the polar ice sheets, our description of the transport of a single species of dissolved impurities should capture the major features of the transport of the overall chemical signal. Our principal goal is to provide a framework for the future analysis and reanalysis of ice-core chemistry.

The interstitial solute concentration c is the mass of impurity dissolved within a unit mass of the premelted liquid. The value of c at a particular depth is determined by equilibrium requirements, which are imposed by contemporary *in situ* conditions, particularly the temperature. Measurements from ice-core samples yield the

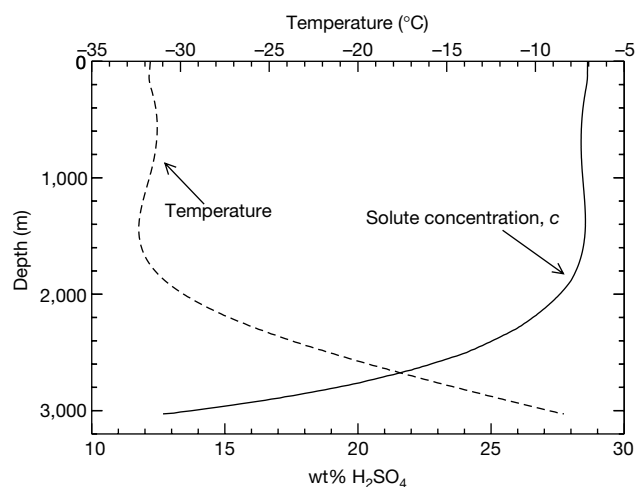


Figure 2 The temperature¹⁴ as a function of depth in the GRIP ice core and the corresponding interstitial concentration of H_2SO_4 . At typical ice-core temperatures, the effects of interfacial curvature have a negligible influence on the total liquid fraction in the veins and nodes^{10,11}, so data from the liquidus relation¹⁶ were used to determine c , after correcting for the pressure dependence of the bulk melting temperature using the Clapeyron equation¹⁹. For example, ref. 16 states that the liquidus concentration at an undercooling of 29.65°C and atmospheric pressure is 28% by weight; at an ice-equivalent depth of 2,000 m, the bulk melting temperature is decreased by about 1.3°C , so we infer that $c \approx 0.28$ when $T \approx -31^{\circ}\text{C}$ at that depth. For comparison, c_b is typically of order 10^{-7} and ϕ is of order 10^{-6} , which is consistent with observations of micrometre-sized veins surrounding grains of a few millimetres in diameter in Antarctic ice¹⁷. There is empirical evidence to support the assumption, at least for the case of H_2SO_4 (refs 17 and 18), that all the soluble impurities are confined to the liquid, but improved data on the locations of impurities within polycrystalline ice should be acquired. Some researchers have referred to solidification experiments²⁰ and assumed that most of the impurities are located within the ice grains. The gradual recrystallization of ice grains over thousands of years occurs slowly enough that there seems to be ample justification for assuming that most impurities remain outside the lattice boundaries. The few recognized exceptions to this rule are F^- , Cl^- and NH_3 (ref. 7). Further analysis shows that the essential behaviour reported here does not change significantly when a finite proportion of the impurities resides within the ice lattice because the lattice impurities are expected to equilibrate with the solute in the veins more rapidly than the solute concentration is changed in response to local temperature variations (see Supplementary Information).

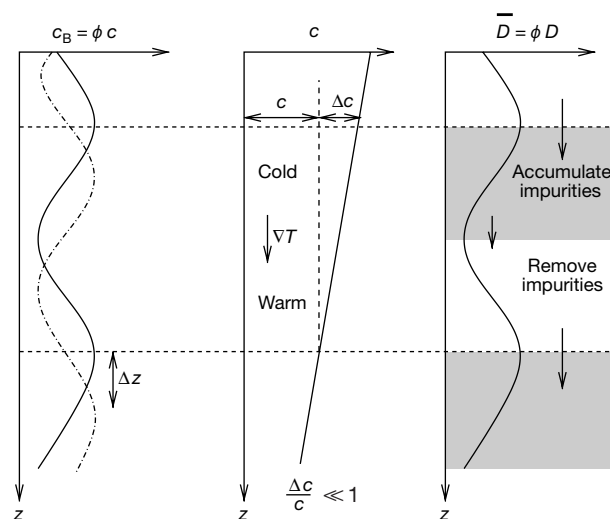


Figure 3 The physical interactions that cause molecular diffusion to alter the c_b -profile in an advective fashion. The left side represents a segment of the c_b -profile that encompasses several annual layers of ice—typically a few centimetres thickness. The corresponding interstitial solute concentration profile $c(z)$ is shown next to this; the temperature contrast across the segment is typically less than 0.01°C so the associated relative change $\Delta c/c$ is small. Hence, the variation in ϕ mirrors the variation in c_b , as does the effective diffusivity $\bar{D} \equiv \phi D$. The arrows on the right indicate that the rate of diffusive solute transport $\bar{D} \nabla c$ is fastest (slowest) where c_b is greatest (smallest). Impurities accumulate and c_b increases in the shaded region, as the rate of transport into that region is greater than the rate of transport out. (In normal diffusion there is no zone of accumulation.) Variations in the transport rate are 90° out-of-phase with variations in the rate of impurity accumulation; this is a characteristic of wave-like behaviour. The solute concentration is fixed by the thermal field, but ϕ adjusts so that the peaks and troughs in the c_b -profile are translated relative to the ice matrix. After a time Δt the bulk concentration profile has moved a distance of $\Delta z = v_c \Delta t$ with its shape unaltered, as shown by the dot-dashed line on the left.

bulk impurity concentration c_B , defined as the mass of impurity contained within a unit mass of the polycrystalline sample. Past climate changes are associated with temporal variations in the rate of impurity deposition so that the value of c_B is often observed to double over a distance of a few centimetres, even in ice that is over 100 kyr old. The *in situ* temperature profile is independent of c_B , however, so the liquidus relation indicates that c experiences far more gradual changes with depth (Fig. 2). The conservation of solute requires that $c_B = \phi c$, where ϕ is the mass fraction of liquid. Because c is relatively constant over short distances, the measured centimetre-scale variations in c_B are mirrored by commensurate fluctuations in ϕ . The solute conservation condition employed in our model predicts that, as the surface energy of curved interfaces acts to make vein radii uniform^{10,11}, variations in c_B must correlate with changes in the total length of veins per unit sample volume. This is confirmed by the observed anti-correlation between c_B and grain size^{21,29,30}.

We are interested in the movement of impurities relative to the ice matrix, which itself moves downward at velocity $\mathbf{v}$ under the influence of gravity. The ice is approximately incompressible and the velocity of the liquid relative to the ice matrix is small, so we can express the solute balance as

$$\frac{\partial(\phi c)}{\partial t} + \mathbf{v} \cdot \nabla(\phi c) = \nabla \cdot (\phi D \nabla c) \quad (1)$$

where D is the diffusion coefficient through the liquid. The left side of equation (1) is the rate of change of the mass of impurity within a volume element that moves with the ice matrix at velocity $\mathbf{v}$. The right side represents the effects of molecular (Fickian) diffusion, which are characterized by the effective diffusivity $\tilde{D} \equiv \phi D$. Equa-

tion (1) can be written in terms of c_B as

$$\frac{\partial c_B}{\partial t} = -(\mathbf{v} + \mathbf{v}_c) \cdot \nabla c_B - (\nabla \cdot \mathbf{v}_c) c_B \quad (2)$$

where

$$\mathbf{v}_c \equiv -\frac{D}{c} \nabla c \quad (3)$$

Moreover, because variations in c_B occur over centimetre-length scales, while variations in $\mathbf{v}_c$ are tied to temperature variations that have characteristic length scales of hundreds of metres, the second (damping) term on the right side of equation (2), involving the divergence of $\mathbf{v}_c$, can be neglected in comparison to the first (advective) term, involving the gradient in c_B (see the Supplementary Information). Equation (2) indicates that, while the ice moves with velocity $\mathbf{v}$, the climate signals contained in the c_B -profile move at a rate that is enhanced by the anomalous velocity $\mathbf{v}_c$. The diffusion of impurities through the liquid alters the c_B -profile in an advective fashion, transporting anomalies in c_B down the solute concentration gradient towards higher temperatures (Fig. 3; see legend for a more detailed explanation of the mechanisms). This transport occurs without significantly altering the amplitudes of the c_B anomalies. This could lead to the misinterpretation of older palaeoclimatic records, which, consistent with the predictions of equation (2), have not undergone the characteristic smoothing that is normally associated with compositional diffusion.

The dashed curve in Fig. 4 shows the magnitude of $\mathbf{v}_c$ as a function of depth for conditions representative of those encountered within the Summit ice cores in Greenland. The anomalous velocity is slow through most of the ice, which is encouraging for the preservation of a consistent age–depth relationship for the long-term averages, measured at the scale of the core bags (55 cm at GRIP), both of soluble constituents that diffuse through the liquid and of palaeoclimatic data that originate from species that move with the local ice velocity. Over tens of thousands of years, however, the motion of the c_B -profile could cause confusion in interpreting the relative timing of palaeoclimatic anomalies. This is illustrated by the solid curve in Fig. 4, which shows a conservative estimate of the separation between the c_B -profile and ice of the same age. At a depth of 2,800 m, which corresponds to the Eemian age in the Summit ice cores, the isochrones are separated by about 50 cm—this is comparable to the thickness of the ice layers that correspond to the sudden cooling events detected in the Eemian ices from the GRIP core, stressing the need for higher-resolution ice-core chemical records.

In order to illuminate the underlying physics of anomalous diffusion, we have presented a simple model based on fundamental mechanisms. In actual ice sheets there are always many species of soluble impurities present in the premelted liquid, and equilibrium with the solid grains is achieved by adjusting ϕ in response to variations in both temperature and solution chemistry. This implies that the variation in c for a particular impurity is not necessarily monotonic in T , and that $\mathbf{v}_c$ varies between different species. In addition, some impurities precipitate onto the grain boundaries at colder temperatures and dissolve only toward the base of the ice sheet, where warmer temperatures make them stable in solution. A more complete understanding of the phase relationships and chemical interactions between the various impurity species is needed before these known effects can be incorporated into predictive models. Chemical interactions might explain why the c_B anomalies are still in the Eemian isotopic events²², for instance, though these observations could also be explained by the hypothesis that the ice stratigraphy has been disrupted²³ in the recent past. The model presented here should be regarded as a test of the potential for diffusion in the unfrozen liquid to disrupt palaeoclimatic records and a challenge to motivate further high-resolution examinations of the deep ice cores. For example, our predictions indicate

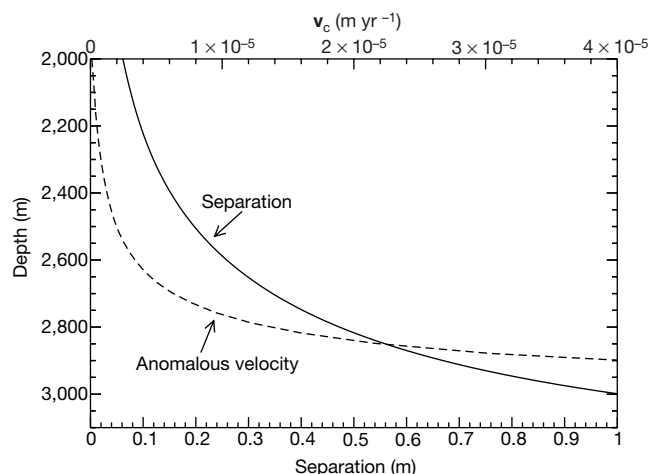


Figure 4 Anomalous diffusion of H_2SO_4 under conditions found at Summit, Greenland. We focus on the bottom portion of the ice sheet where the decrease in c (Fig. 2) causes $\mathbf{v}_c$ to become significant (see equation (3)). The dashed curve shows $|\mathbf{v}_c|$ as a function of depth when the contemporary temperature field at GRIP is used to infer the dependence of c on depth. For solute transport through the veins of polycrystalline ice, D is reduced to one-third the normal molecular diffusivity when the veins are randomly oriented with respect to the concentration gradient¹⁰; as a lower bound we take $D \approx 5 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ (ref. 16). The solid curve shows the separation between the c_B -profile and ice of the same age. We used a steady-state Dansgaard–Johnsen²⁷ flow model for the ice with a kink point at 600 m above the base (20% of the 3,000-m ice thickness) and an accumulation rate of 0.23 m yr^{-1} . Past changes in the temperature profile were investigated using time-dependent heat and mass-flow models; in the depth range shown, these lead to variations in predicted magnitudes of $\mathbf{v}_c$ of up to 30% over the past 100 kyr. Accumulation rates during the last glacial period were lower than today by as much as a factor of four²⁸ so our steady-state flow model underestimates the age of the ice near the base. As this is where $|\mathbf{v}_c|$ is highest, the data presented here should tend to underestimate the separation between isochrones of the ice and isochrones of the c_B -record.

that anomalous diffusion can displace the sulphuric acid spikes that are used as stratigraphic markers to correlate timescales between different ice cores and the tephra deposits contained in other sedimentary records²⁴. Our model explains how diffusion preserves the amplitudes of anomalies in the c_B record, but the anomalies themselves are translated relative to the surrounding ice. Efforts should be made to account for this behaviour when analysing data from the older portions of ice cores by increasing spatial resolution. This could be particularly important when the relative timing of concentration peaks is needed to test theories for the causal links between the various climate proxies. □

Received 2 October 2000; accepted 2 April 2001.

- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Alley, R. B. Ice-core evidence of abrupt climate changes. *Proc. Natl Acad. Sci. USA* **97**, 1331–1334 (2000).
- Dash, J. G., Fu, H. Y. & Wettlaufer, J. S. The premelting of ice and its environmental consequences. *Rep. Prog. Phys.* **58**, 115–167 (1995).
- Wettlaufer, J. S. Ice surfaces: Macroscopic effects of microscopic structure. *Phil. Trans. R. Soc. Lond. A* **357**, 3403–3425 (1999).
- Paterson, W. S. B. *The Physics of Glaciers* 3rd edn, 8–25 (Pergamon, Oxford, 1994).
- Wolff, E. W. in *Chemical Exchange Between the Atmosphere and Polar Ice* (eds Wolff, E. W. & Bales, R. C.) 541–560 (NATO ASI Series I, Vol. 43, Springer, Berlin, 1996).
- Ramseier, R. O. Self-diffusion of tritium in natural and synthetic ice monocrystals. *J. Appl. Phys.* **38**, 2553–2556 (1967).
- Nye, J. F. in *Physics and Chemistry of Ice* (eds Maeno, N. & Hondoh, T.) 200–205 (Hokkaido Univ. Press, 1992).
- Nye, J. F. The geometry of water veins and nodes in polycrystalline ice. *J. Glac.* **35**, 17–22 (1989).
- Mader, H. M. Observations of the water-vein system in polycrystalline ice. *J. Glac.* **38**, 333–347 (1992).
- Nye, J. F. Diffusion of isotopes in the annual layers of ice sheets. *J. Glac.* **44**, 467–468 (1998).
- Johnsen, S. J. *et al.* in *Ice Physics and the Natural Environment* (eds Wettlaufer, J. S., Dash, J. G. & Untersteiner, N.) 89–107 (NATO ASI Series I, Vol. 56, Springer, Berlin, 1999).
- Gundestrup, N. S., Dahl-Jensen, D., Johnsen, S. J. & Rossi, A. Bore-hole survey at dome GRIP—1991. *Cold Reg. Sci. Technol.* **21**, 399–402 (1993).
- Cuffey, K. M. *et al.* Large Arctic temperature-change at the Wisconsin–Holocene glacial transition. *Science* **270**, 455–458 (1995).
- Weast, R. C. (ed.) *CRC Handbook of Chemistry and Physics* 68th edn, D219–D269 (CRC Press, Boca Raton, 1987).
- Mulvaney, R., Wolff, E. W. & Oates, K. Sulphuric acid at grain-boundaries in Antarctic ice. *Nature* **331**, 247–249 (1988).
- Fukazawa, H., Sugiyama, K., Mae, S. J., Narita, H. & Hondoh, T. Acid ions at triple junction of Antarctic ice observed by Raman scattering. *Geophys. Res. Lett.* **25**, 2845–2848 (1998).
- Wood, S. E. & Battino, R. *Thermodynamics of Chemical Systems* (Cambridge Univ. Press, Cambridge, 1990).
- Gross, G. W., Chen-ho, W., Bryant, L. & McKee, C. Concentration dependent solute redistribution at the ice/water phase boundary. II. Experimental investigation. *J. Chem. Phys.* **62**, 3085–3092 (1975).
- Thorsteinsson, T., Kipfstuhl, J., Eicken, H., Johnsen, S. J. & Fuhrer, K. Crystal size variations in Eemian-age ice from the GRIP ice core, central Greenland. *Earth Planet. Sci. Lett.* **131**, 381–394 (1995).
- Steffensen, J. P., Clausen, H. B., Hammer, C. U., Legrand, M. & De Angelis, M. The chemical composition of cold events within the Eemian section of the Greenland Ice Core Project ice core from Summit. *J. Geophys. Res.* **102**, 26747–26754 (1997).
- Alley, R. B. *et al.* Comparison of deep ice cores. *Nature* **373**, 393–394 (1995).
- Zielinski, G. A. Use of paleo-records in determining variability within the volcanism–climate system. *Quat. Sci. Rev.* **19**, 417–438 (2000).
- Smith, C. S. Grains, phases and interfaces: An introduction to microstructure. *Trans. Metall. Soc. AIME* **175**, 15–51 (1948).
- Ohtomo, M. & Wakahama, G. Growth-rate of recrystallization in ice. *J. Phys. Chem.* **87**, 4139–4142 (1983).
- Dansgaard, W. & Johnsen, S. J. A flow model and a time scale for the ice core from Camp Century, Greenland. *J. Glac.* **8**, 215–223 (1969).
- Cuffey, K. M. & Clow, G. D. Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition. *J. Geophys. Res.* **102**, 26383–26396 (1997).
- Thorsteinsson, T., Kipfstuhl, J. & Miller, H. Textures and fabrics in the GRIP ice core. *J. Geophys. Res.* **102**, 26583–26599 (1997).
- De la Chapelle, S., Castelnau, O., Lipenkov, V. & Duval, P. Dynamic recrystallization and texture development in ice as revealed by the study of deep ice cores in Antarctica and Greenland. *J. Geophys. Res.* **103**, 5091–5105 (1998).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We acknowledge R. Alley, J. G. Dash, D. P. Winebrenner, S. G. Warren, G. W. Gross, S. F. Johnsen, J. F. Nye, E. J. Steig, J. P. Steffensen and E. W. Wolff for discussions that have influenced this work. We also thank H. M. Mader for providing the photograph for Figure 1b. Support for this research has been provided by the US National Science Foundation.

Correspondence and requests for materials should be addressed to A.W.R. (e-mail: awrempe@apl.washington.edu).

The post-spinel transformation in Mg_2SiO_4 and its relation to the 660-km seismic discontinuity

Sang-Heon Shim*, Thomas S. Duffy* & Guoyin Shen†

* Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA

† CARS, University of Chicago, Chicago, Illinois 60637, USA

The 660-km seismic discontinuity in the Earth's mantle has long been identified with the transformation of $(\text{Mg,Fe})_2\text{SiO}_4$ from γ -spinel (ringwoodite) to $(\text{Mg,Fe})\text{SiO}_3$ -perovskite and $(\text{Mg,Fe})\text{O}$ -magnesiowüstite. This has been based on experimental studies of materials quenched from high pressure and temperature^{1–3}, which have shown that the transformation is consistent with the seismically observed sharpness and the depth of the discontinuity at expected mantle temperatures⁴. But the first *in situ* examination of this phase transformation in Mg_2SiO_4 using a multi-anvil press⁵ indicated that the transformation occurs at a pressure about 2 GPa lower than previously thought (equivalent to ~600 km depth) and hence that it may not be associated with the 660-km discontinuity. Here we report the results of an *in situ* study of Mg_2SiO_4 at pressures of 20–36 GPa using a combination of double-sided laser-heating and synchrotron X-ray diffraction in a diamond-anvil cell. The phase transformation from γ - Mg_2SiO_4 to MgSiO_3 -perovskite and MgO (periclase) is readily observed in both the forward and reverse directions. In contrast to the *in situ* multi-

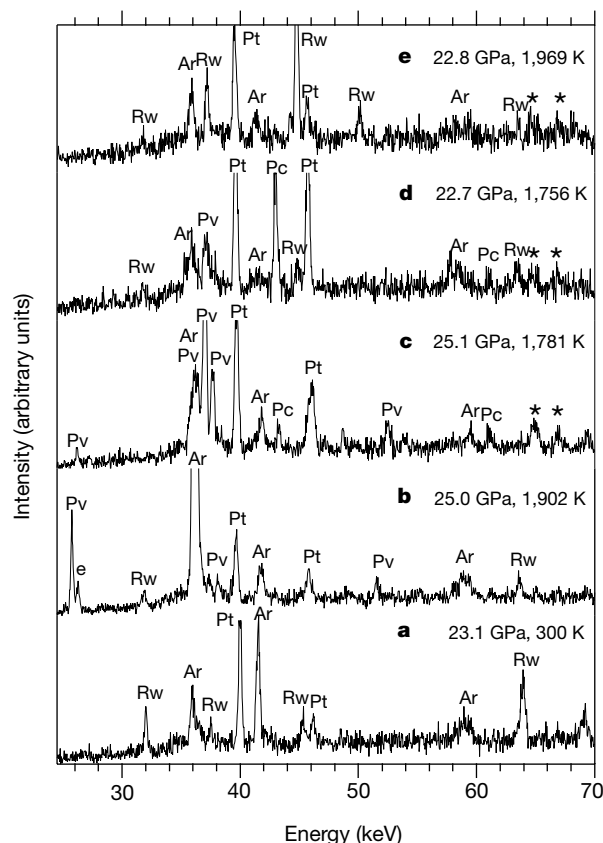


Figure 1 a–e, Representative X-ray diffraction patterns at the indicated P – T conditions. Peak identifications are: Rw, ringwoodite; Pv, perovskite; Pc, periclase; Pt, platinum; Ar, argon; asterisk, platinum fluorescence; e, detector escape peak.

anvil-press study⁵, we find that the pressure and temperature of the post-spinel transformation in Mg_2SiO_4 is consistent with seismic observations^{4,6} for the 660-km discontinuity.

Pure synthetic forsterite was mixed with 10 wt% platinum which serves as an internal pressure scale and laser absorber. A 15- μm -thick foil of the sample mixture was loaded in a 100- μm hole in a steel gasket and compressed by 300- μm diamond anvils. Argon served as a pressure transmitting and insulation medium. Ruby chips were also loaded in an unheated area.

The laser-heating system at the GSECARS sector of the Advanced Photon Source provides a radially homogeneous temperature profile over 20–30 μm by using a TEM₀₁ Nd:YLF laser mode, and reduces the axial thermal gradient by heating from both sides^{7,8}. Temperature profiles were measured using imaging spectrometers on each side. Temperature was determined by fitting the radiation spectra, corrected for system response, to Planck's equation. A three-dimensional averaging technique was used to calculate the mean temperature over the volume exposed to X-rays⁷. Including radial and axial gradients, temperature fluctuations during X-ray exposure, and the fitting residual, the temperature uncertainty (1σ) was ± 50 –150 K in these experiments.

Energy dispersive X-ray diffraction measurements were performed using a small (5 $\mu\text{m} \times 7 \mu\text{m}$) horizontally focused X-ray beam and a solid-state detector (Fig. 1). We oscillated the diamond cell ($\pm 20^\circ$) about its loading axis to minimize the effect of preferred orientation. The primary pressure scale used here is the pressure–volume–temperature (P – V – T) equation of state (EOS) of platinum⁹. Pressure was also measured using the ruby scale¹⁰ before and after heating.

Below 20 GPa, we performed two heating runs to synthesize ringwoodite. The pressure was then increased to 20.4 GPa and a third run was performed. We observed an increase of 2–4 GPa during heating due to the thermal pressure^{11,12}. After 20 minutes of heating, the periclase (200) peak appeared and co-existed with ringwoodite diffraction lines. By increasing the temperature from 1,672(50) K to 1,942(75) K (numbers in parentheses are 1σ uncertainties), we observed a pressure decrease from 24.6(4) GPa to 22.3(6) GPa as a result of thermal relaxation¹³ (see P – T path for run 3 in Fig. 2). At this point, the periclase (200) line disappeared, owing to back transformation to the lower-pressure phase. In the

fourth run, we increased pressure to 23.1 GPa (Fig. 1a). After eight minutes of heating, diagnostic perovskite lines, (002)+(110) and (004)+(220) doublets, appeared (Fig. 1b). The periclase (200) line was also observed in other patterns during this run. We performed two more runs at 25–29 GPa (run 5 in Fig. 2) and 30–35 GPa, where we observed complete transformation from ringwoodite to perovskite+periclase (Fig. 1c). To observe the transformation from perovskite+periclase to ringwoodite, we decreased pressure and performed three heating runs at 27–30, 25–26 and 21–25 GPa. For the last run, ringwoodite appeared after 6 minutes of heating and the transformation was complete after 24 minutes (Fig. 1d, e). Additional details regarding the P – T paths are provided in the Supplementary Information.

We observe that the low- and high-pressure phase assemblages coexist within a range of about 2 GPa (Fig. 2). This scatter is due to both temperature uncertainty and kinetic effects. The temperature uncertainty is 50–150 K, and this propagates to a pressure uncertainty of 0.3–0.9 GPa. Kinetics and the P – T path may also be important. For example, as shown above in the P – T path for run 3 (Fig. 2), we observed that the high-pressure assemblage synthesized above the phase boundary can survive at slightly below the phase boundary for 5 minutes before complete back transformation.

Using data points for which we observed a mixture of low- and high-pressure phases, we obtained the phase boundary. We fixed the slope of the phase boundary to the average from two multi-anvil studies^{2,5} (-2.75 MPa K^{-1}), because our data scatter precludes constraining the slope reliably. A weight was assigned to each data point using its pressure and temperature uncertainties. The transformation pressure was found to be $23.7 \pm 1.1 \text{ GPa}$ at 1,800 K. However, the boundary obtained in ref. 5 is 2.6 GPa lower than this, which is a statistically significant difference at the 2σ level (Fig. 2).

In order to examine this result, we investigated several error sources. In the laser-heated diamond cell, the greatest error source is the temperature uncertainty¹⁴. The error in temperature propagates to pressure through the thermal-pressure term. The sensitivity of pressure for temperature, $(\partial P/\partial T)_P$, is calculated for platinum using its P – V – T EOS⁹. To have a 2.6-GPa error, the temperature must be overestimated by 380 K. However, including all random error sources, our temperature uncertainty is less than 150 K. Systematic error sources have also been examined, and are estimated to be less than 100 K for this heating system⁸. Using lattice strain theory¹⁵,

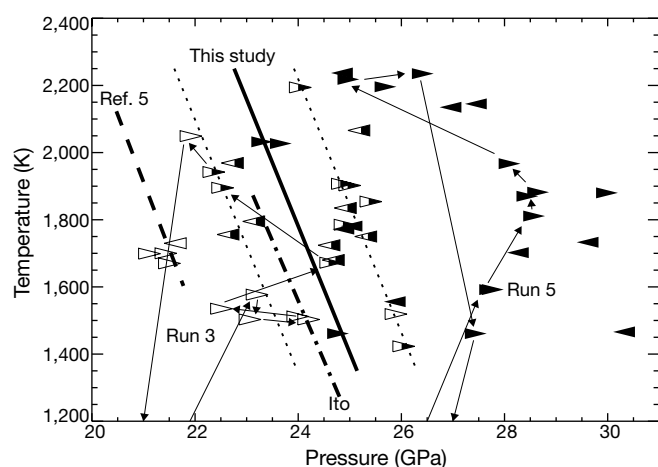


Figure 2 Phase diagram of Mg_2SiO_4 . Right-pointing and left-pointing triangles represent respectively data points obtained during compression and decompression. The presence of ringwoodite is shown by open symbols and that of perovskite+periclase is shown by solid symbols. Half-filled symbols represent coexistence of ringwoodite and perovskite+periclase. Two representative P – T paths during heating runs (run 3 and 5) are shown by thin solid lines with arrowheads. The heavy dash-dotted and dashed lines show phase boundaries from previous studies^{2,5}. The phase boundary obtained here is shown by the heavy solid line with 1σ uncertainty (thin dotted lines).

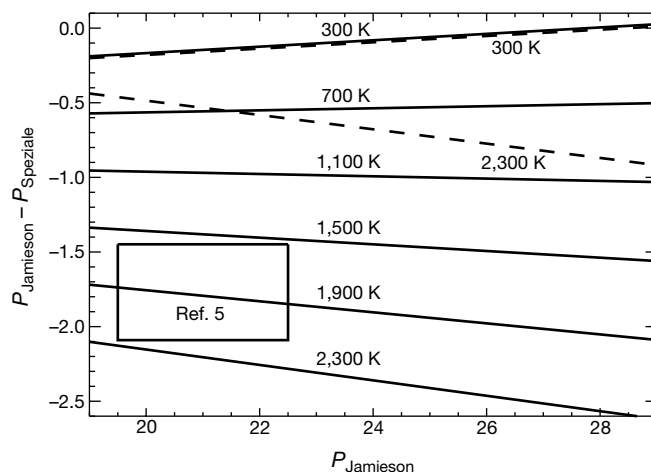


Figure 3 Pressure difference between two equations of state for periclase^{16,20} at different P – T conditions. The P – T range in the experiments of ref. 5 is indicated by the rectangle. The effect of changing the Grüneisen parameter in the EOS of Speziale *et al.*¹⁶ to the value used by Jamieson *et al.*²⁰ is shown by dashed lines. See text for details. P_{Jamieson} and P_{Speziale} are the pressures calculated using the MgO EOS from refs 20 and 16, respectively.

differential stresses were also investigated and found to be negligible at high temperature.

We also compared pressures obtained from different materials in our sample, such as periclase¹⁶ and ringwoodite¹⁷. On average, the platinum pressure scale⁹ overestimates pressure by 0.2 GPa compared to periclase and by 2.0 GPa compared to ringwoodite. We also examined the difference between platinum and ruby pressures before and after heating. On average, platinum yields a larger pressure by ~ 1.2 GPa. However, the ruby chips were located at the edge of the sample chamber where the pressure should be lower, owing to the radial gradient. Thus, except for ringwoodite, all pressures are internally consistent within 1.2 GPa.

One possibility for the discrepancy between our result and those of Irifune *et al.*⁵ is the inconsistency of the equations of state of platinum and gold. Irifune *et al.* justified the use of the gold pressure scale of ref. 18 on the basis of an earlier study¹⁹ that found good agreement between the gold EOS of Anderson *et al.*¹⁸ and the periclase EOS by Jamieson *et al.*²⁰. In our study, we observed that the platinum pressure scale by Holmes *et al.*⁹ shows good agreement with the periclase pressure scale by Speziale *et al.*¹⁶. Thus by comparing the periclase EOS values, we can obtain an inter-comparison of gold and platinum. Figure 3 shows the difference between the periclase EOS of Jamieson *et al.*²⁰ and Speziale *et al.*¹⁶. We observe a maximum difference of 2.1 GPa between the two periclase pressure calculations at P - T conditions of the post-spinel transformation. This means that the gold EOS¹⁸ yields ~ 2 GPa lower pressure than the platinum EOS⁹. We also observed a 2 GPa underestimation of pressure by the ringwoodite EOS¹⁷, which was also obtained using the gold pressure scale of ref. 18.

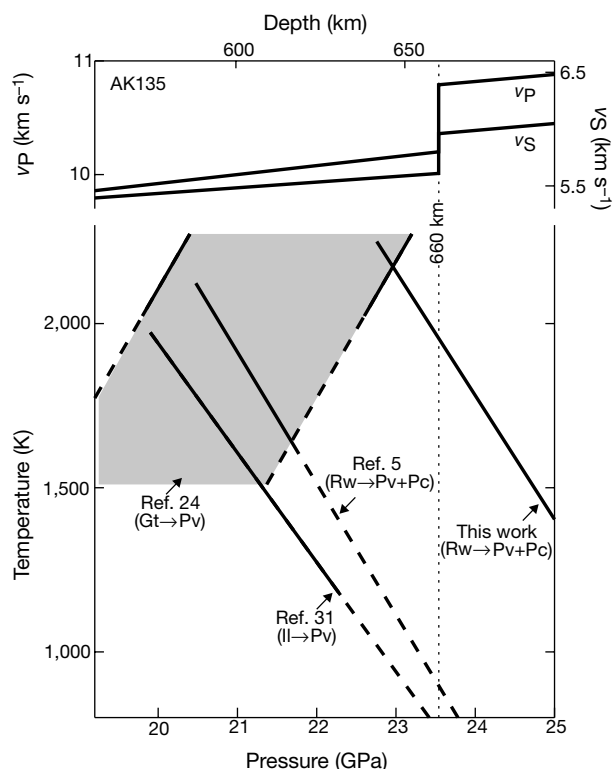


Figure 4 Seismic velocity profiles⁶ and phase boundaries of mantle constituents near the 660-km discontinuity. Top, P-wave and S-wave velocities as a function of depth in the Earth. Data from earth model AK135 (ref. 6). Bottom, phase boundaries of mantle constituents (solid lines, dashed where extrapolated). The post-ilmenite (Il $\rightarrow$ Pv, where Il indicates ilmenite and Pv indicates perovskite) boundary is from ref. 31. The post-garnet (Gt $\rightarrow$ Pv) transformation with 6 mol% Al_2O_3 (from ref. 24) is plotted as a shaded area where majorite garnet (Gt) and perovskite coexist. The post-spinel (Rw $\rightarrow$ Pv+Pc, where Rw indicates ringwoodite and Pc indicates periclase) boundaries are from ref. 5 and this study.

The discrepancy between the two periclase scales is mainly caused by the use of different Grüneisen parameters. As shown in Fig. 3, the discrepancy decreases drastically just by changing the Grüneisen parameter used by Speziale *et al.*¹⁶ (1.52 ± 0.05) to the value used by Jamieson *et al.*²⁰ (1.32). The former workers obtained the Grüneisen parameter using a thermodynamic relationship, $\gamma = \alpha K_S / \rho C_P$ (α is the thermal expansivity, K_S is the adiabatic bulk modulus, ρ is the density and C_P is the isobaric heat capacity), and accurately determined thermodynamic parameters, whereas the value used by the latter workers was based on low-accuracy estimates from porous shock wave data. In addition, the EOS of Speziale *et al.*¹⁶ satisfies a wide range of high P - T data including static compression^{21,22} and shock wave data²³.

A recent study²⁴ that intercompared pressure standards suggested that the gold and platinum pressure scales are consistent. However, this study also found that various different standards yield pressures that differ by ~ 2 GPa at 24 GPa. We also note that Anderson later concluded that the gold EOS may not be reliable²⁵. These results emphasize the need for further studies of pressure standards. Other possible sources of systematic error—such as the effect of pressure on thermocouple e.m.f. in multi-anvil studies²⁶, temperature gradients, deviatoric stress, and stress inhomogeneity—also need to be investigated.

Phase boundaries for mantle minerals determined using *in situ* techniques at pressures near 660-km depth are shown in Fig. 4. Apart from this work, all determinations were performed in the multi-anvil press using the gold scale of Anderson *et al.*¹⁸. The post-spinel and post-ilmenite boundaries yield low temperatures (< 900 K) at 660-km depth. However, a variety of geophysical constraints including xenoliths, basalt melting temperatures, and *in situ* data for the α - β transformation in Mg_2SiO_4 suggest that transition zone temperatures are $\sim 1,600$ – $1,900$ K (refs 27–29). At these temperatures, the post-spinel boundary would occur at ~ 610 -km depth. Because no discontinuity is observed near here, this would place severe limits on the olivine abundance of the transition zone and further require that the 660-km discontinuity be associated with a chemical change or the garnet–perovskite transformation. However, seismological data suggest the 660-km discontinuity does not prevent the penetration of subducting slabs³⁰. Furthermore, to associate the post-garnet boundary²⁴ with the 660-km discontinuity, aluminium enrichment relative to pyrolite is required. Although the thickness and depth of this boundary will be affected by bulk composition²⁴, the slope of post-garnet boundary conflicts with the seismically observed negative Clapeyron slope of the 660-km discontinuity⁴. In contrast, our observations of the post-spinel transformation are fully consistent with geophysical data for the 660-km discontinuity.

We have thus shown that the laser-heated diamond-anvil cell can accurately determine phase boundaries of mantle constituents. We find that the post-spinel phase boundary lies within a plausible mantle temperature range at 660 km depth, and thus this transformation remains the best candidate to explain this seismic discontinuity. □

Received 25 October 2000; accepted 6 April 2001.

1. Liu, L.-G. Post-oxide phases of forsterite and enstatite. *Geophys. Res. Lett.* **2**, 417–419 (1975).
2. Ito, E. & Takahashi, E. Postspinel transformations in the system Mg_2SiO_4 – Fe_2SiO_4 and some geophysical implications. *J. Geophys. Res.* **94**, 10637–10646 (1989).
3. Ito, E., Akaogi, M., Topor, L. & Navrotsky, A. Negative pressure-temperature slopes for reactions forming MgSiO_3 perovskite from calorimetry. *Science* **249**, 1275–1278 (1990).
4. Helffrich, G. Topography of the transition zone seismic discontinuities. *Rev. Geophys.* **38**, 141–158 (2000).
5. Irifune, T. *et al.* The postspinel phase boundary in Mg_2SiO_4 determined by *in situ* X-ray diffraction. *Science* **279**, 1698–1700 (1998).
6. Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from travel times. *Geophys. J. Int.* **122**, 108–124 (1995).
7. Shim, S.-H., Duffy, T. S. & Shen, G. The stability and P - V - T equation of state for CaSiO_3 perovskite in the earth's lower mantle. *J. Geophys. Res.* **105**, 25955–25968 (2000).
8. Shen, G., Rivers, M. L., Wang, Y. & Sutton, S. R. Laser heated diamond cell system at the Advanced Photon Source for *in situ* x-ray measurements at high pressure and temperature. *Rev. Sci. Instrum.* **72**, 1273–1282 (2001).

9. Holmes, N. C., Moriarty, J. A., Gathers, G. R. & Nellis, W. J. The equation of state of platinum to 660 GPa (6.6 Mbar). *J. Appl. Phys.* **66**, 2962–2967 (1989).
10. Mao, H.-K., Xu, J. & Bell, P. M. Calibration of the ruby pressure gauge to 800 kbar under quasihydrostatic conditions. *J. Geophys. Res.* **91**, 4673–4676 (1986).
11. Heinz, D. L. Thermal pressure in the laser-heated diamond anvil cell. *Geophys. Res. Lett.* **17**, 1161–1164 (1990).
12. Fiquet, G. *et al.* X-ray diffraction of periclase in a laser-heated diamond-anvil cell. *Phys. Earth Planet. Inter.* **95**, 1–17 (1996).
13. Kavner, A. & Duffy, T. S. Pressure-volume-temperature paths in the laser-heated diamond anvil cell. *J. Appl. Phys.* **89**, 1907–1914 (2001).
14. Boehler, R. High-pressure experiments and the phase diagram of lower mantle and core materials. *Rev. Geophys.* **38**, 221–245 (2000).
15. Singh, A. K. The lattice strains in a specimen (cubic system) compressed nonhydrostatically in an opposed anvil device. *J. Appl. Phys.* **73**, 4278–4286 (1993).
16. Speziale, S. *et al.* Quasi-hydrostatic compression of magnesium oxide to 52 GPa: implications for the pressure–volume–temperature equations of state. *J. Geophys. Res.* **106**, 515–528 (2001).
17. Meng, Y. *et al.* In situ high P–T X-ray diffraction studies on three polymorphs (α , β , γ) of Mg_2SiO_4 . *J. Geophys. Res.* **98**, 22199–22207 (1993).
18. Anderson, O. L., Isaak, D. G. & Yamamoto, S. Anharmonicity and the equation of state for gold. *J. Appl. Phys.* **65**, 1534–1543 (1989).
19. Funamori, N. *et al.* Thermoelastic properties of MgSiO_3 perovskite determined by in situ X-ray observations up to 30 GPa and 2000 K. *J. Geophys. Res.* **101**, 8257–8269 (1996).
20. Jamieson, J. C., Frits, J. N. & Manghnani, M. H. in *High-Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) 27–48 (Center for Academic Publications, Tokyo, 1982).
21. Fei, Y. Effect of temperature and composition on the bulk modulus of $(\text{Mg},\text{Fe})\text{O}$. *Am. Mineral.* **84**, 272–276 (1999).
22. Dewaele, A., Fiquet, G., Andrault, D. & Häusermann, D. P–V–T equation of state of periclase from synchrotron radiation measurements. *J. Geophys. Res.* **105**, 2869–2877 (2000).
23. Svendsen, B. & Ahrens, T. J. Shock-induced temperatures of MgO . *Geophys. J. R. Astron. Soc.* **91**, 667–691 (1987).
24. Hirose, K. *et al.* In situ measurements of the phase transition boundary in $\text{Mg}_3\text{Al}_2\text{Si}_2\text{O}_{12}$: implications for the nature of the seismic discontinuities in the Earth's mantle. *Earth Planet. Sci. Lett.* **184**, 567–573 (2001).
25. Anderson, O. L. The volume dependence of thermal pressure in perovskite and other minerals. *Phys. Earth Planet. Inter.* **112**, 267–283 (1999).
26. Rubie, D. C. Characterising the sample environment in multianvil high-pressure experiments. *Phase Transit.* **68**, 431–451 (1999).
27. Jackson, I. & Ridgen, S. M. in *The Earth's Mantle: Composition, Structure, and Evolution* (ed. Jackson, I.) 405–460 (Cambridge University Press, New York, 1998).
28. Vacher, P., Mocquet, A. & Sotin, C. Computation of seismic profiles from mineral physics: the importance of the non-olivine components for explaining the 660 km depth discontinuity. *Phys. Earth Planet. Inter.* **106**, 275–298 (1998).
29. Morishima, H. *et al.* The phase boundary between α - Mg_2SiO_4 and β - Mg_2SiO_4 determined by in-situ X-ray observation. *Science* **265**, 1202–1203 (1994).
30. van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
31. Kuroda, K. *et al.* Determination of the phase boundary between ilmenite and perovskite in MgSiO_3 by in situ X-ray diffraction and quench experiments. *Phys. Chem. Mineral.* **27**, 523–532 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Speziale for experimental assistance, and F. Dahlen, T. Irifune, Y. Fei, O. Anderson and I. Jackson for discussions. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to S.-H.S. (e-mail: sangshim@princeton.edu).

High-pressure polymorphs of olivine and the 660-km seismic discontinuity

L. Chudinovskikh & R. Boehler

Max-Planck Institut für Chemie, Postfach 3060, 55020 Mainz, Germany

It had long been accepted that the 400-km seismic discontinuity in the Earth's mantle results from the phase transition of $(\text{Mg},\text{Fe})_2\text{SiO}_4$ -olivine to its high-pressure polymorph β -spinel (wadsleyite), and that the 660-km discontinuity results from the breakdown of the higher-pressure polymorph γ -spinel (ringwoodite) to MgSiO_3 -perovskite and $(\text{Mg},\text{Fe})\text{O}$ -magnesiowüstite^{1–4}. An *in situ* multi-anvil-press X-ray study⁵ indicated, however, that the phase boundary of the latter transition occurs at pressures 2 GPa lower than had been found in earlier studies using multi-anvil

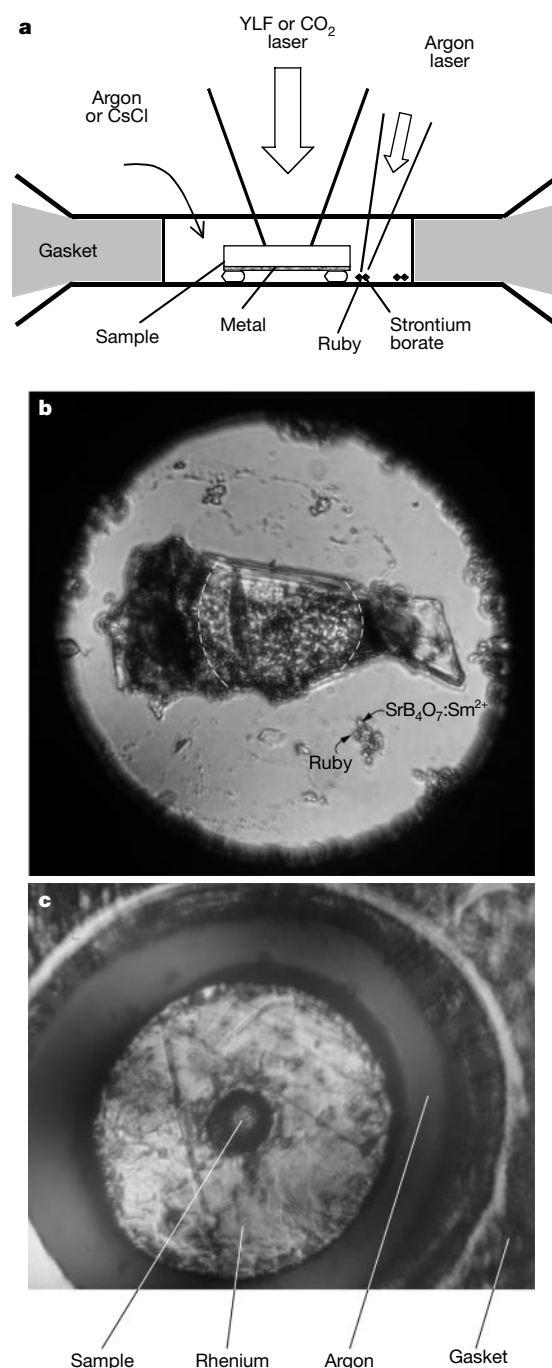


Figure 1 Diamond-cell arrangements used in this study to heat samples of Mg_2SiO_4 -forsterite. **a**, Schematic cross-section. Samples were polished disks of pure, synthetic Mg_2SiO_4 -forsterite, 10–15 μm thick and with diameters of 20 to 100 μm , thermally insulated from the bottom diamond by grains of forsterite. Argon or CsCl were used as pressure media. The pressure chambers were typically 170–180 μm in diameter and 45–55 μm in thickness. The samples were either heated directly with a 150 W CO_2 laser ($\lambda = 10.5 \mu\text{m}$), or were sputtered with films of Mo, Re or Cr ($<1 \mu\text{m}$ thick) and then heated with a 50 W YLF (yttrium-lithium-fluoride) laser ($\lambda = 1.05 \mu\text{m}$)—and in some experiments with the combined beams of an additional 20 W YLF laser. **b**, Molybdenum-sputtered forsterite sample after heating. The heated portion showed full conversion to γ - Mg_2SiO_4 . Thermal pressure increase was measured from two adjacent grains of ruby and strontium borate excited with an argon laser. **c**, Small crystal (15–20 μm in size) inside a Re microfurnace with outer diameter 80 μm , inner diameter 20 μm , and a thickness of 30 μm . This furnace was uniformly heated with the YLF laser in an argon pressure medium. This heating method essentially eliminated temperature gradients in the sample.

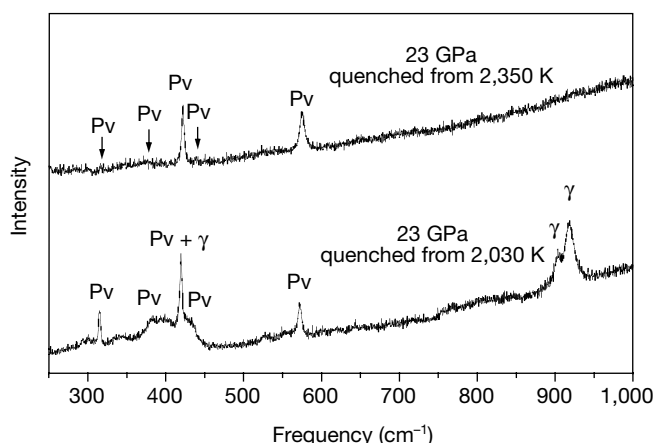


Figure 2 Examples of phase identification by Raman spectroscopy. Top spectrum, run no. 6; bottom spectrum, run no. 9-1; see Table 1. Peak positions for γ - Mg_2SiO_4 and perovskite (Pv) are from refs 29 and 11, respectively.

recovery experiments⁶ and laser-heated diamond-anvil cells⁷. Such a lower-pressure phase boundary would be irreconcilable with the accuracy of seismic measurements of the 660-km discontinuity, and would thus require a mineral composition of the mantle that is significantly different from what is currently thought. Here, however, we present measurements made with a laser-heated diamond-anvil cell which indicate that γ - Mg_2SiO_4 is stable up to pressure and temperature conditions equivalent to 660-km depth in the Earth's mantle (24 GPa and 1,900 K) and then breaks down into MgSiO_3 -perovskite and MgO (periclase). We paid special attention to pressure accuracy and thermal pressure in our experiments, and to ensuring that our experiments were performed under nearly hydrostatic, inert pressure conditions using a variety of heating methods. We infer that these factors are responsible for the different results obtained in our experiments compared to the *in situ* multi-anvil-press study⁵.

Pressures within the Earth are well constrained from geophysical data, but temperatures have to be derived from measurements of phase boundaries of the relevant high-pressure minerals. Such measurements require careful calibration of pressure and temperature. Both multi-anvil work⁸ and laser-heating experiments in the diamond cell⁷ on the transition from $(\text{Mg}_{0.88}\text{Fe}_{0.12})_2\text{SiO}_4$ -olivine to

the β -phase suggest that for an olivine-dominated mantle the temperature at a depth of 400 km is about 1,800 K. Since the discovery of MgSiO_3 -perovskite from one of the early laser-heating experiments⁹, the phase boundary between γ - Mg_2SiO_4 and perovskite+ MgO has been measured in multi-anvil presses⁶ and laser-heated diamond cells⁷. The importance of this phase boundary for estimating the temperature at 660 km depth lies in the fact that it seems not to be significantly affected by iron contents of up to 20% or whether the low-pressure phase is of Mg_2SiO_4 or MgSiO_3 composition⁶. Thus, temperature estimates are insensitive to the choice of mineralogical models for the mantle. From the diamond-cell work⁷ and the slope of the phase boundary⁶, a temperature of $1,900 \pm 100$ K was derived at 24 GPa, the pressure at 660 km depth. Measurements of pressure and temperature in the diamond cell use the ruby scale and Planck's radiation law, respectively. In multi-anvil presses, pressures are obtained by extrapolating from lower-pressure phase transitions and temperatures are measured from thermocouples. Taking into account the marked differences in the experimental techniques, the direct measurements of the phase boundaries of both major transitions using diamond-cell or multi-anvil methods yielded remarkable agreement.

In contrast, recent measurements in a multi-anvil press by synchrotron X-ray diffraction⁵ suggested a lowering of the γ - Mg_2SiO_4 to perovskite transition by at least 2 GPa, thus making this transition unsuitable to explain the 660-km discontinuity. In that study, two possibilities were considered as explanations of this discrepancy: the first is a reconsideration of the mineral composition of the mantle, and the second addresses the reliability of the equation of state of the internal pressure calibrants. Nevertheless, this estimated phase boundary was recently used to calibrate measurements of the MgSiO_3 -ilmenite to perovskite transition¹⁰ which, consequently, resulted in similar low transition pressures when compared to previous measurements⁶ and to an estimate of that phase boundary from calorimetric data and spectroscopic entropy measurements in the diamond cell¹¹.

Pressure measurements in the multi-anvil X-ray press rely on a high-temperature equation of state of a calibrant added to the sample. However, gold, which was used in the work reported in ref. 5, shows variations of $\pm 15\%$ (refs 12–14) in its bulk modulus as measured by X-ray diffraction, even at room temperature. It is very unlikely that the theoretically derived high-temperature equation of state of gold¹⁵ used in ref. 5 is more accurate than these measurements, and so the accuracy of the pressures reported in ref. 5 is therefore less than claimed. Thus, uncertainties in the equations of

Table 1 Results of phase transitions in Mg_2SiO_4

Run no.	Sample	Pressure medium	P (GPa)*	T (K)†	Laser type	Heating duration (min)	Run products‡
6	Forsterite (pure)	CsCl	23.0	2,350	CO_2	10	Pv
4	Forsterite (Mo)	Ar	22.0	1,950	YLF	10	γ
4-1	Forsterite (Mo)	Ar	24.0	1,950	YLF	20	Pv
9	Forsterite (Mo)	Ar	21.1	2,050	YLF	10	γ
9-1	Forsterite (Mo)	Ar	23.0	2,030	YLF	15	$\gamma + \text{Pv}$
13-2	Forsterite (Re)	CsCl	22.2	2,230	YLF	20	γ
13-3	Forsterite (Re)	CsCl	22.2	2,030	CO_2	27	γ
13-4	Forsterite (Re)	CsCl	22.5	2,200	CO_2	22	$\gamma + \text{Pv}$
13-7	Forsterite (Re)	CsCl	22.0	2,870	CO_2	5	Pv
18-1	Forsterite (Re)	Ar	25.8	1,800	YLF+	45	Pv
18-6	Forsterite (Re)	Ar	23.0	1,880	YLF+	30	Pv
18-7	Forsterite (Re)	Ar	22.6	1,800	YLF+	30	Pv + γ
11	Forsterite in Re-ring	Ar	21.3	2,100	YLF	30	$\beta + \gamma$
12	Forsterite in Re-ring	Ar	21.2	2,000	YLF	20	$\beta + \gamma$
15-7	Forsterite (Cr)	Ar	21.7	2,350	CO_2	10	Pv
15-8	Forsterite (Cr)	Ar	20.0	2,300	CO_2	10	β , no Pv
15-9	Forsterite (Cr)	Ar	21.1	2,360	CO_2	10	β
15-10	Forsterite (Cr)	Ar	21.9	2,430	CO_2	10	β
15-11	Forsterite (Cr)	Ar	22.7	2,330	CO_2	15	β
15-12	Forsterite (Cr)	Ar	24.0	2,350	CO_2	15	Pv

See Fig. 1 legend for experimental details. † Increasing pressure; ‡ decreasing pressure. YLF, yttrium-lithium-fluoride. Pv, perovskite.

* Uncorrected pressure, ± 0.2 GPa, average from various ruby chips distributed throughout the pressure chamber.

† Temperature, ± 200 K.

‡ From Raman spectra.

state used in ref. 5 may easily explain a large part of the discrepancy in the measured phase boundaries. Another source of error—which is widely ignored in multi-anvil work—lies in the temperature measurement. It has been shown that the effect of pressure on the e.m.f.s of thermocouples is substantial, even at much lower pressures and temperatures¹⁶. But the correction is complex, and depends on the individual experimental conditions. Even though it may be substantial at pressures exceeding 20 GPa, it is unlikely, however, to be the sole source of the discrepancy in the perovskite phase boundary.

Because of the large discrepancy for this geophysically important phase boundary between laser-heated diamond cells and the recent X-ray diffraction measurements in multi-anvil presses, we measured the phase transitions in the Mg_2SiO_4 system in the laser-heated diamond cell using a variety of methods (Fig. 1a–c)—we paid special attention to the reliability of the pressure measurements. Experimental details of the lasers, laser stability, and temperature measurements have been reported elsewhere^{7,17}. Heating durations of up to 30 min were usually sufficient for partial or complete phase transformation. After temperature quenching, which happened in less than a millisecond, the high-pressure phases of the laser-heated samples could be unambiguously identified by Raman spectroscopy (Fig. 2). For the Cr-sputtered samples, additional evidence for the presence of MgO was obtained from the fluorescence line of $\text{MgO}:\text{Cr}^{3+}$ (ref. 18). Two reversal experiments (numbers 18 and 15 in Table 1) were performed on fully converted perovskite–magnesiowüstite assemblages by lowering the pressure in between heating cycles. In one experiment, the γ -phase was identified at 23.4 GPa after the sample had been heated for 30 min at 1,800 K, and in the second experiment the β -phase formed at 20.8 GPa and 2,300 K.

The crucial element in measuring the phase boundary is the accurate determination of the pressure. In a diamond cell, soft, nearly hydrostatic, pressure media such as noble gases can be used, and pressures can be measured reliably by the well calibrated ruby fluorescence method¹⁹. The absolute accuracy of the room-temperature ruby scale has been tested²⁰ in MgO by measuring

ruby fluorescence at the same time as measuring elasticity (by Brillouin spectroscopy) and measuring volume (by X-ray diffraction). It has been shown that pressure gradients in the pressure chamber are significantly reduced during laser-heating of a sample in an argon pressure medium²¹. Thus, thermal pressure increases in the pressure chamber during heating may be approximated from rubies at the unheated edge of the pressure chamber. In the present experiments, ruby fluorescence shifts equivalent to 1.0 ± 0.2 GPa were typically observed at pressures around 22 GPa and sample temperatures around 2,000 K. The actual pressure increase was probably lower, because any temperature increase in the ruby causes a strong red shift in the fluorescence spectrum^{22,23}—as does increasing pressure. We performed additional measurements of the thermal pressure closer to the heated sample by the simultaneous measurement of the fluorescence shifts from adjacent chips of ruby ($\text{Al}_2\text{O}_3:\text{Cr}^{3+}$) and strontium borate ($\text{SrB}_4\text{O}_7:\text{Sm}^{2+}$), 3–5 μm in size (Fig. 1b). At room temperature the pressure shifts of both materials are equally well calibrated^{19,23–25}. In contrast to the large temperature shift of the ruby fluorescence line^{22,23}, the temperature shift of strontium borate is nearly zero at ambient and slightly elevated pressures^{24,26}, thus probably providing a more accurate pressure measurement in a hot environment. As an example, the thermal pressure in the cell geometry shown in Fig. 1b, as calculated from the fluorescence shift of strontium borate of 0.16 nm, was 0.6 GPa for a sample temperature of 2,000 K. This thermal-pressure increase was checked using the known pressure and temperature shifts of ruby. The width of the ruby fluorescence line with increasing temperature has been measured up to 6 GPa (refs 22, 23). From these data, we estimated the temperature of the ruby chip and the portion of the fluorescence shift measured during heating that was due to this temperature increase. From the remaining shift the thermal pressure increase was then estimated as 0.8 GPa, which is close to the value obtained from the shift of the adjacent strontium borate chip. Thus, pressures reported in this study measured at room temperature after heating were probably between 0.6 and 1.0 GPa higher during heating.

Temperatures in the laser-heated diamond cell are measured spectroscopically using Planck's law. This method requires only one assumption about the wavelength dependence on emissivity—which is probably negligible for iron-free minerals in the wavelength range of our measurements. Temperature uncertainties in the present study are about ± 200 K. Temperature gradients in the sample are kept to a minimum by using thin samples surrounded by pressure media of low thermal conductivity, thereby allowing a defocused laser beam. Typical temperature profiles in silicate or metallic samples have been reported previously^{27,28}.

We performed 20 experiments to determine the phase boundaries between β - and γ - Mg_2SiO_4 and the MgSiO_3 -perovskite+MgO assemblage, and the run conditions and results are listed in Table 1. The pressures in this table are uncorrected, measured at room temperature after heating, and are averages measured from ruby chips randomly distributed between sample and cell edge. The maximum variation in these pressures was ± 0.2 GPa. Temperature fluctuations and the precision of the measurements were ± 100 K. Figure 3 shows the data after applying a thermal pressure correction of +0.8 GPa, the average of the present thermal-pressure estimates. γ - Mg_2SiO_4 (ringwoodite) forms at pressures that are between 2 and 3 GPa higher than the suggested⁵ γ - Mg_2SiO_4 -perovskite phase boundary, and the present data are in good agreement with the phase boundary between γ - Mg_2SiO_4 and the MgSiO_3 -perovskite+MgO assemblage found earlier^{6,7}. At higher temperatures, the present phase-transition data between β - Mg_2SiO_4 and MgSiO_3 -perovskite+MgO, not previously measured by multi-anvil techniques, is in accordance with both the extrapolated γ - Mg_2SiO_4 to β - Mg_2SiO_4 transition as measured in ref. 8, and the γ - Mg_2SiO_4 to MgSiO_3 -perovskite+MgO phase boundary.

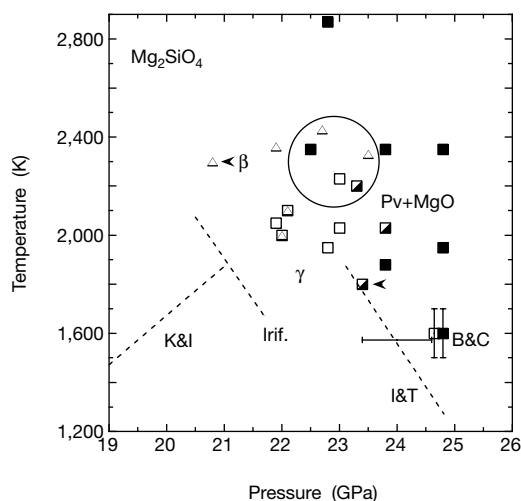


Figure 3 Phase diagram of Mg_2SiO_4 . Triangles are β -phase, open squares are γ -phase, and solid squares are MgSiO_3 -perovskite. Mixed symbols represent mixed run products. The pressures listed in Table 1 were corrected for thermal pressure by +0.8 GPa. Previous estimates of the γ - β transition are from Katsura and Ito⁸ (K&I); previous estimates of the γ to perovskite+MgO transition are from Ito and Takahashi⁶ (I&T), Bohler and Chopelas⁷ (B&C), and Irifune *et al.*⁵ (Irif.). The horizontal and vertical bars represent the reported uncertainties in pressure and temperature, respectively. Reversal experiments are indicated by arrows. The circle represents pressure and temperature conditions for which all three phases have been identified in different runs.

The circle shown in Fig. 3 shows an area where all three phases have been identified in different runs. The centre of this circle is in excellent agreement with the γ - β -perovskite+MgO triple point reported²⁹ at 22.9 GPa and 2,260 K.

The present data strongly support earlier estimates of the phase-boundary temperature between ringwoodite and perovskite of about 1,900 K at 24 GPa, the pressure at a depth of 660 km in the Earth's mantle. The measurements also show that phase relations in simple systems can be measured more reliably in laser-heated diamond cells than by using multi-anvil methods; this is because measurements of pressure and temperature are more accurate, and the method provides the capability to use an inert and nearly hydrostatic pressure environment. Moreover, phase boundaries can be measured over much larger temperature ranges. Additional constraints on the phase diagram will have to come from calorimetric and spectroscopic measurements—as shown for the MgSiO₃ high-pressure polymorphs¹¹. Larger samples and multi-anvil devices will be required for more complex systems, but pressure estimates will need to be improved by the systematic investigation of high-temperature equations of state of inert pressure calibrants such as gold. These could be provided by synchrotron X-ray measurements in laser-heated diamond cells. □

Received 7 December 2000; accepted 28 March 2001.

1. Ringwood, A. E. Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta* **55**, 2083–2110 (1991).
2. Poirier, J. P. in *Introduction to the Physics of the Earth's Interior* (eds Putnis, A. & Liebermann, R. C.) 214–226 (Cambridge Univ. Press, Cambridge, 1991).
3. Zhao, Y. & Anderson, D. L. Mineral physics constraints on the chemical composition of the Earth's lower mantle. *Phys. Earth Planet. Inter.* **85**, 273–292 (1994).
4. Navrotsky, A. *Physics and Chemistry of Earth Materials* (Cambridge Univ. Press, Cambridge, 1994).
5. Irfune, T. E. A. The postspinel phase boundary in Mg₂SiO₄ determined by in situ X-ray diffraction. *Science* **279**, 1698–1700 (1998).
6. Ito, E. & Takahashi, E. Postspinel transformations in the system Mg₂SiO₄-Fe₂SiO₄ and some geophysical implications. *J. Geophys. Res.* **94**, 10637–10646 (1989).
7. Boehler, R. & Chopelas, A. A new approach to laser heating in high pressure mineral physics. *Geophys. Res. Lett.* **18**, 1147–1150 (1991).
8. Katsura, T. & Ito, E. The system Mg₂SiO₄-Fe₂SiO₄ at high pressures and temperatures: precise determination of stabilities of olivine, modified spinel, and spinel. *J. Geophys. Res.* **94**, 15663–15670 (1989).
9. Liu, L. A new high pressure phase of spinel. *Earth Planet. Sci. Lett.* **41**, 398–404 (1978).
10. Kuroda, K. et al. Determination of the phase boundary between ilmenite and perovskite in MgSiO₃ by in situ X-ray diffraction and quench experiments. *Phys. Chem. Miner.* **27**, 523–532 (2000).
11. Chopelas, A. Estimates of mantle relevant Clapeyron slopes in the MgSiO₃ system from high-pressure spectroscopic data. *Am. Miner.* **84**, 233–244 (1999).
12. Manghnani, M. H. et al. in *Physics of Solids under High Pressure* (ed. Schilling, J. S.) 47–55 (North Holland, Amsterdam, 1981).
13. Heinz, D. L. & Jeanloz, R. The equation of state of the gold calibration standard. *J. Appl. Phys.* **55**, 885–893 (1984).
14. Ming, L. C., Xiong, D. & Manghnani, M. H. Isothermal compression of Au and Al to 20 GPa. *Physica B* **139**, 174–176 (1986).
15. Anderson, O. L., Isaak, D. G. & Yamamoto, S. Anharmonicity and the equation of state for gold. *J. Appl. Phys.* **65**, 1534–1543 (1989).
16. Getting, I. C. & Kennedy, G. C. Effect of pressure on the emf of chromel-alumel and platinum-platinum 10% rhodium thermocouples. *J. Appl. Phys.* **41**, 4552–4562 (1970).
17. Boehler, R. High-pressure experiments and the phase diagram of lower mantle and core materials. *Rev. Geophys.* **38**, 221–245 (2000).
18. Chopelas, A. The fluorescence sideband method for obtaining acoustic velocities at high compressions: application to MgO and MgAl₂O₄. *Phys. Chem. Miner.* **23**, 25–37 (1996).
19. Mao, H. K., Xu, J. & Bell, P. M. Calibration of the ruby pressure gauge to 800 kbar under quasi-hydrostatic conditions. *J. Geophys. Res.* **91**, 4673–4676 (1986).
20. Zha, C. S., Mao, H. K. & Hemley, R. J. A primary pressure scale to 55 GPa from new measurements of equation of state and elasticity for MgO. *AIRAPT Proc.* **136**, (1999).
21. Serghiou, G., Zerr, A., Chudinovskikh, L. & Boehler, R. The coesite-stishovite transition in a laser-heated diamond cell. *Geophys. Res. Lett.* **22**, 441–444 (1995).
22. Shimomura, O., Yamaoka, S., Nakazawa, H. & Fukunaga, O. in *High Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) 49–60 (Center for Academic Publishing of Japan, Tokyo, 1982).
23. Ragan, D. D., Gustavsen, R. & Schiferl, D. Calibration of the ruby R1 and R2 fluorescence shifts as a function of temperature from 0 to 600 K. *J. Appl. Phys.* **72**, 5539–5544 (1992).
24. Datchi, F., Le Toullec, R. & Loubeyre, P. Improved calibration of the SrB₃O₇:Sm²⁺ optical pressure gauge: advantages at very high pressures and high temperatures. *J. Appl. Phys.* **81**, 3333–3339 (1997).
25. Yen, J. & Nicol, M. Temperature dependence of the ruby luminescence method for measuring high pressures. *J. Appl. Phys.* **72**, 5535–5538 (1992).
26. Lacam, A. & Chateau, C. The SrB₃O₇:Sm²⁺ optical sensor for diamond anvil cells. *J. Appl. Phys.* **66**, 366–370 (1989).
27. Boehler, R. & Zerr, A. Perovskite temperature profile. *Science* **265**, 723 (1994).
28. Boehler, R. Melting and thermal expansion of iron in uniformly laser-heated diamond anvil cells. *High Pressure Res.* **5**, 702–704 (1990).

29. Chopelas, A., Boehler, R. & Ko, T. Thermodynamics and behavior of γ -Mg₂SiO₄ at high pressure: implications for Mg₂SiO₄ phase equilibrium. *Phys. Chem. Mineral.* **21**, 352–359 (1994).

Acknowledgements

We thank A. Zerr for assistance with the CO₂-laser heating experiments.

Correspondence and requests for materials should be addressed to R.B. (e-mail: boe@mpch-mainz.mpg.de).

Ecological and evolutionary processes at expanding range margins

C. D. Thomas*, E. J. Bodsworth*, R. J. Wilson*, A. D. Simmonds*, Z. G. Davies*, M. Musche*† & L. Conrad†

* Centre for Biodiversity and Conservation, School of Biology, University of Leeds, Leeds LS2 9JT, UK

Many animals are regarded as relatively sedentary and specialized in marginal parts of their geographical distributions^{1,2}. They are expected to be slow at colonizing new habitats. Despite this, the cool margins of many species' distributions have expanded rapidly in association with recent climate warming^{3–10}. We examined four insect species that have expanded their geographical ranges in Britain over the past 20 years. Here we report that two butterfly species have increased the variety of habitat types that they can colonize, and that two bush cricket species show increased fractions of longer-winged (dispersive) individuals in recently founded populations. Both ecological and evolutionary processes are probably responsible for these changes. Increased habitat breadth and dispersal tendencies have resulted in about 3- to 15-fold increases in expansion rates, allowing these insects to cross habitat disjunctions that would have represented major or complete barriers to dispersal before the expansions started. The emergence of dispersive phenotypes will increase the speed at which species invade new environments, and probably underlies the responses of many species to both past¹¹ and future climate change.

Our first example concerns the silver-spotted skipper butterfly *Hesperia comma* in southern England. Near the coldest edges of their geographical ranges, ectothermic animals are often restricted to unusually warm habitats, such as areas of short vegetation on sheltered, south-facing (in the northern temperate zone) hillsides^{1,2}. If they are limited by temperature, such populations should respond to climate warming by expanding into nearby habitats that were previously too cool. In 1982, *H. comma* was largely restricted to south- and south-west facing chalk grassland fragments in south-east England¹². By 2000, however, it had colonized a wider range of aspects, including east-, west- and north-facing hillsides (Kuiper's test, $n = 82$, $m = 228$, $k = 5,478$, $P < 0.001$, where n is the number of habitat units occupied in 1982, m is the number colonized between 1982 and 2000, and k is the test statistic; Fig. 1a, b). High population densities were found on all aspects in 2000. Using these and additional habitat criteria¹² (see Methods), we were able to estimate the distributions of habitat in the eastern half of the South Downs hills (about 400 km² of south-east England), corresponding to 1982 and 2000 definitions of habitat. We estimated that 105 thermally

† Present address: Luther-Universität-Wittenberg, Institut für Zoologie, Domplatz 4, 060108 Halle, Germany.

Table 1 Latitudinal patterns of distribution of *A. agestis*

Latitudinal band (OS 10-km grid)	Number of squares occupied by <i>A. agestis</i>		Proportion of <i>A. agestis</i> squares within <i>H. chamaecistus</i> distribution*		Proportion (number) of UK squares occupied by <i>H. chamaecistus</i>
	1970–1982	1995–1999	1970–1982	1995–1999	
1–10 (south)	46	56	0.43	0.41	0.23 (27)
11–20	188	278	0.65	0.59	0.47 (171)
21–30	80	254	0.66	0.39	0.28 (119)
31–40	19	93	0.74	0.41	0.22 (79)
41–50 (north)	0	10	na	0.30	0.33 (68)

* The association between *A. agestis* and *H. chamaecistus* distributions (at 10-km grid resolution) becomes weaker between 1970–1982 and 1995–1999, but more so in northern (latitude 21–40N) than in southern (1–20N) Britain (four-way interaction term of log-linear model: latitude $\times$ year $\times$ butterfly presence $\times$ *H. chamaecistus* presence; $\chi^2 = 14.01$, d.f. = 1, $P = 0.0002$).

suitable habitat patches (total habitat area 2.00 km²) would be available with the 1982 definition of habitat, compared with 175 patches (3.92 km²) with the 2000 definition of the same type of vegetation on any aspect.

Expansion rates are likely to increase with habitat availability. For *H. comma* at its northern limit, there are now more patches to colonize, shorter distances between them, and more populations generating emigrants. We used Hanski's spatially realistic incidence function (metapopulation) model^{13–15} to simulate expansion away from the only two patches of habitat occupied by *H. comma* in the South Downs in 1982. This simulation was done in two ways, using patch networks on the basis of the 1982 and 2000 definitions of habitat suitability. Distance expanded was estimated as the mean distance from the main patch occupied in 1982 to the ten furthest populations (Fig. 1c; distances to the nearby, extremely small 1982 population were very similar). We first estimated parameters for the model in a different patch network (Surrey) occupied by *H. comma*, elsewhere in southern England, before using these parameter estimates to make independent predictions of expansion in the South Downs network. In 18 years (generations), the observed distribution of *H. comma* reached 16.37 km (Fig. 1c), corresponding well to the mean 14.35 km (s.d. 1.79) modelled expansion using the 2000 habitat network (Fig. 1e). In contrast, the observed expansion fell well outside the 99.9% confidence limit of simulations using the 1982 network (Fig. 1d; mean 4.98 km, s.d. 0.80). An approximate doubling of habitat availability led to an approximate trebling in expansion rate.

The second example considers the brown argus butterfly (*Aricia agestis*), which has also expanded both its habitat and geographical range over the past 20 years. Between 1970–1982 (ref. 16) and 1995–1999 (ref. 6), this butterfly increased 1.2-fold in the south of England, and 4.9-fold further north, in central England (Table 1). In 1970–1982, the northern part of the butterfly's distribution was mainly restricted to the distribution of one host plant *Helianthemum chamaecistus* (Cistaceae), a plant itself restricted to fragmented chalk and limestone grasslands^{16,17}. Further south, there was a greater tendency for *A. agestis* populations to use *Geranium* and *Erodium* species (Geraniaceae) (Table 1). These plants are more widespread than *H. chamaecistus* and occur in both chalk and non-chalk habitats¹⁷. Between 1970–1982 and 1995–1999, *A. agestis*

expanded disproportionately into Geraniaceae-containing areas—especially *Geranium*—in the northern half of its distribution (Table 1; $P = 0.0002$). *H. chamaecistus* often grows on sheltered or southerly facing hillsides that are likely to be warmer than most *Geranium*-containing habitats: we propose that most northern *Geranium* habitats were too cool to achieve population growth in the 1960s and 1970s, but warm enough in the 1980s and 1990s.

The disproportionate expansion of habitat breadth in the northern half of the distribution could be explained completely by ecological processes. However, once habitat range has begun to expand for ecological reasons, phenotypes able to use either a wide range of habitats or the commonest type of habitat available in marginal areas may show relatively fast rates of range expansion. In *A. agestis*, expanding and non-expanding marginal populations differed in choice of host plant, and thereby habitat. We carried out choice experiments in the field, during which wild butterflies were experimentally exposed to four potential host plant species. Long-established and non-expanding marginal populations of *A. agestis* from limestone and sand dune habitats chose to lay eggs on host plants appropriate to their natural habitats (Table 2; 92% of eggs on *Geranium molle* and *Erodium cicutarium* at Gwithian and Mexico Towans sand dunes in Cornwall; 74% of eggs on *H. chamaecistus* plants on limestone in the Creuddyn Peninsula, north Wales). In contrast, females from populations that had been established recently chose to lay two-thirds of their eggs on *G. molle*, irrespective of whether they came from limestone (Barnack, north Cambridgeshire, colonized in 1994; C. Gardener, personal communication) or sand dune habitats (Gibraltar Point, Lincolnshire, colonized in 1994; K. Wilson, personal communication) (Table 2). Females in these populations chose to lay eggs on the most widespread host plant used during range expansion (*G. molle*), rather than on the host plant that was used naturally in the habitats where the populations occurred (*H. chamaecistus* at Barnack; *E. cicutarium* at Gibraltar Point). Nonetheless, both populations retained the ability to lay on *H. chamaecistus* (Table 2).

The data are compatible with a genetic contribution to host plant choice, but breeding experiments are required to confirm this. Butterflies from the Cornish and North Wales populations that had been reared in a common environment (laboratory) were released in a field containing experimental *H. chamaecistus* and

Table 2 Percentage of eggs laid on plants by four *A. agestis* populations in choice experiments

Population	Natural hosts	Expansion*	<i>Helianthemum chamaecistus</i>	<i>Geranium molle</i>	<i>Geranium dissectum</i>	<i>Erodium cicutarium</i>	n (eggs)
North Wales	<i>H. chamaecistus</i>	40%	74.1	14.8	7.4	3.7	27
Cornwall	Geraniaceae	0%	2.0	56.0	6.0	36.0	50
Barnack	<i>H. chamaecistus</i>	1,600%	12.9	70.2	14.0	2.9	183
Gibraltar Point	<i>E. cicutarium</i> †	250%	14.1	63.7	14.3	7.9	1,360

Schreiber–Ray–Hare tests: overall population $\times$ host plant interaction: SS/MS_{total} = 23.0, d.f. = 9, $P = 0.006$. Population $\times$ host plant interactions were significant for all pairwise comparisons between populations except Barnack versus Gibraltar Point ($P = 0.688$) and Cornwall versus Gibraltar point ($P = 0.093$). χ^2 tests: four populations and four host plants, $\chi^2 = 143.8$, d.f. = 9, $P < 0.0001$. All pairwise comparisons between populations were significant at $P < 0.0001$, except Barnack versus Gibraltar Point ($P = 0.042$).

* Per cent increase between 1970–1982 and 1995–1999 in 10-km grid squares occupied in 50-km squares centred on each site.

† Some *Geranium* species may also be used, but this has not yet been observed.

G. molle plants. The released females differed in choice of host plant species in the expected direction, consistent with a genetic interpretation (140 eggs from the Cornish butterflies on *G. molle*, 25 on *H. chamaecistus*; 168 eggs from the north Wales butterflies on *G. molle*, 189 on *H. chamaecistus*; $\chi^2 = 66.6$, degrees of freedom (d.f.) = 1, $P < 0.001$). Females reared in the laboratory from the same populations showed no detectable effects of larval environment (host plant species or temperature) on subsequent egg-laying preferences (E.J.B., unpublished observations). Phenotypic carry-over (maternal) effects could affect these single-generation results, but could not plausibly explain the results of the choice experiments at Barnack and Gibraltar Point. Virtually all Barnack butterflies developed as larvae on *H. chamaecistus* and virtually all Gibraltar Point butterflies developed on *E. cicutarium*, and must have done so for the previous 9–10 generations. However, given a choice, both populations laid eggs disproportionately on *G. molle*. Similarly, learning by adults (or conditioning on emergence) could not be responsible for the observed patterns of host choice. Butterflies with similar adult experiences of habitat and host plants differed strongly

in choice of plants (for example, north Wales and Barnack), whereas adults exposed to dissimilar habitats and host plants exhibited similar host choice characteristics (Table 2; Barnack and Gibraltar Point).

Host choice phenotypes in the expanding region may have arisen from selection within each population ancestry during range expansion, or the entire range expansion may have been initiated from populations that already possessed the host choice characteristics of Barnack and Gibraltar Point. Whatever the origin, expansion into *Geranium*-containing habitats in the north enabled the butterfly repeatedly to bridge gaps of greater than 14 km in the fragmented *H. chamaecistus* distribution^{6,16,17}. This is a distance that *A. agestis*^{18,19} and comparable butterflies^{20,21} would be extremely unlikely to cross in one step, without the availability of stepping-stones of other types of habitat.

Range expansion by flying insects may also select for increased flight ability^{22–24}. Our third example concerns two species of bush crickets that exhibit adult wing polymorphisms. Both species have been spreading northwards and inland from distributions formerly confined to specific habitats in southern, coastal areas (Fig. 2)^{7–9}. The long-winged cone-head *Conocephalus discolor* has two forms: long-winged and extra-long-winged (macropterous). Many populations established in the past 20 years show higher frequencies of extra-long-winged individuals, with a low (and fairly constant) frequency of this form in populations established greater than 20 years ago (Fig. 2c). Roesel's bush cricket *Metrioptera roeselii* has a short-winged form that cannot fly and a long-winged form that can. This species also shows increased frequencies of the more dispersive form in populations that have recently been established (Fig. 2d).

Environmental variables, such as temperature, population density and photoperiod, are known to affect whether cricket nymphs will mature to become long- or short-winged adults²⁵. Thus, increased frequencies of longer wing morphologies in recently founded populations could represent (1) purely plastic responses to these new environments; or (2) genetic differences affecting, for example, the population density or temperature that a nymph must experience before it will develop into a longer-winged adult. We have not established the genetic basis for the observed differences among populations; however, plastic responses to new environments are unlikely to provide the whole explanation. Sites that have been colonized recently tend to have relatively low population densities and would therefore be expected to possess low frequencies of longer-winged forms²⁵, the opposite of the pattern observed. Variation in temperature (longer-winged morphologies expected in hotter regions) and latitude (equivalent to photoperiod) are also unlikely to explain the differences, given comparable results for the two species that have different spatial patterns of invasion (Fig. 2). The observed patterns may reflect the increasing number of colonization bottlenecks experienced by younger populations, and/or selection against dispersal in populations once they have been established. The latter is probable if there is a trade-off between investment in flight-related and reproductive structures²⁵.

Regardless of the proportional contributions of plastic versus genetic change to the observed patterns, the implications for range expansion are profound. If we assume that virtually all long-distance movements are achieved by the longer-winged morphologies (certainly the case for *M. roeselii*), this represents approximately 4-fold and 14-fold increases in long-distance dispersal for *C. discolor* and *M. roeselii*, respectively (based on fractions of longer-winged individuals, pooled across populations established less than 10 versus more than 20 years ago).

Improving environmental conditions at existing margins, in this instance regional warming at cool margins, are likely to initiate range extensions purely on the basis of ecological, physiological and population-dynamic processes—requiring no evolutionary change. Once an expansion is initiated, individuals and populations that expand most rapidly are likely to be favoured, and expanding range

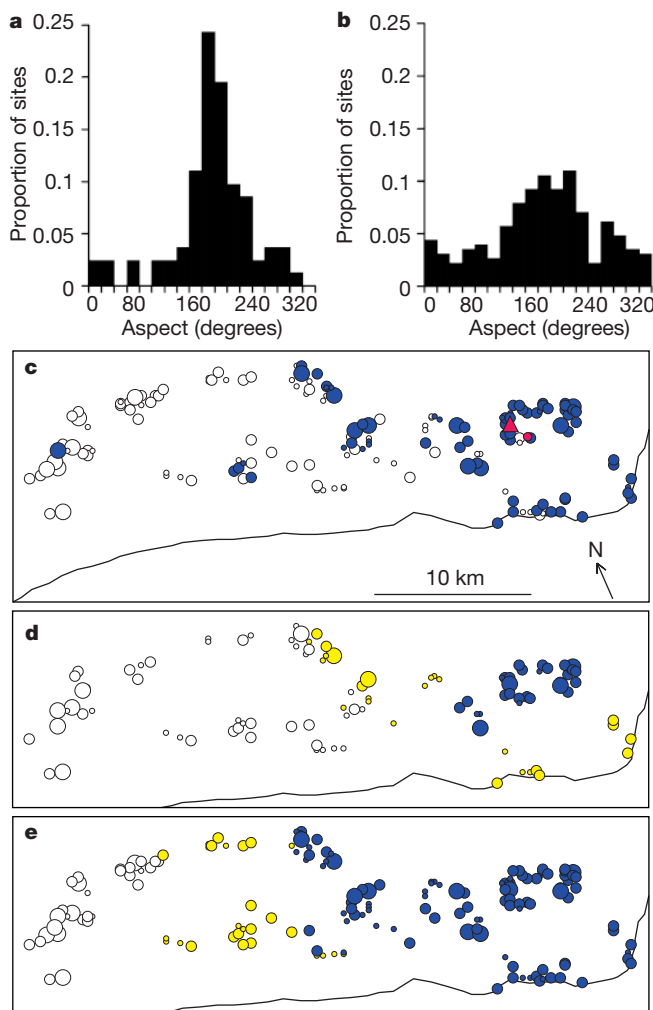


Figure 1 Habitat and range expansion by *H. comma*. **a**, **b**, Aspects of populated *H. comma* habitats in 1982 (**a**), and colonized habitats between 1983–2000 (**b**). **c**, Range of *H. comma* in the South Downs in 1982 (red; triangle contains > 95% of 1982 individuals) and in 2000 (blue and red). Unoccupied habitat (open circles) in 2000 is for the 2000 habitat network. **d**, Range of *H. comma* simulated with 1982 habitat network (blue circles, populated in 2000 in $\geq 50\%$ of simulation runs; yellow circles, < 50% of simulation runs; and open circles, never colonized). **e**, Range of *H. comma* simulated with 2000 habitat network. Symbol sizes exaggerate patch areas (ha); small < 0.5, medium 0.5 to < 5, large ≥ 5 . Line shows south coast of England; Beachy Head is at the south-eastern point.

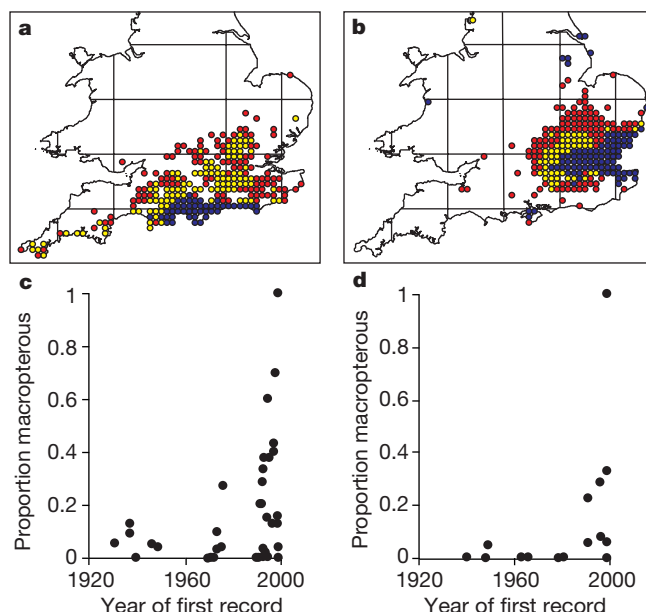


Figure 2 Expanding distributions of *C. discolor* (a) and *M. roeselii*^{7–9} (b). Lines show national 100-km grid squares. Circles show 10-km records: blue, recorded between 1961–1987; yellow, first recorded between 1988–1996; red, first recorded between 1997–1999. c, d, Proportions of macropterous individuals in populations sampled in

2000, according to date of first record for *C. discolor* (c) (proportion extra-long-winged; Spearman correlation $r = 0.349$, $n = 40$, $P = 0.027$), and *M. roeselii* (d) (proportion long-winged; Spearman correlation $r = 0.698$, $n = 17$, $P = 0.002$).

fronts may become characterized by dispersive generalists or by specialists on habitats common in the area of expansion. Once most habitats in a region have been colonized, specialists (in different habitats) and less dispersive forms may be favoured again, especially if there is a trade-off between dispersal and reproduction^{25,26}. Evolutionary stasis during range changes has been emphasized^{27,28}, but transient evolutionary changes in habitat choice, flight behaviour and wing morphology require specific and detailed studies to detect. Such changes may be typical of expanding range fronts. Putative examples include increased frequencies of long-winged forms of ground beetles invading Canada²², increased thorax sizes (containing flight muscles) of butterfly species expanding their ranges in Europe^{23,24}, biased orientation of Africanized honey bee swarms invading Mexico²⁹, and the development of increasingly dispersive seeds during the post-glacial expansion of lodgepole pine in North America¹¹. These changes may dramatically increase rates of range expansion.

Our results should not be interpreted as evidence that all species will be able to change their geographical ranges rapidly in response to recent and predicted climate change. Extremely large numbers of relatively sedentary and specialized species may fail to initiate any expansion across human-modified landscapes, and some expansion is required before the selective pressures described here can result in the evolution of increased habitat range and migration rate. □

Methods

Hesperia comma

Hesperia comma adults and eggs (on the grass *Festuca ovina*) were mapped from July to September 2000. All habitat type was surveyed within 20 km of the 1982 distribution¹², and within 10–15 km of the 2000 distribution, using accepted patch definitions³⁰. Comparison of aspects is based on sites (sub-sites where aspects varied) in the Chiltern Hills, North Downs and South Downs. The 1982 network was defined as sparse chalk grassland containing 'ideal' *F. ovina* tufts growing in slight hollows, next to bare ground¹² with aspects 100–300° (containing greater than 90% 1982 populations in south east England). The 2000 network had the same vegetation/*F. ovina* characteristics, but included grassland fragments of all aspects (the weak bias in Fig. 1b arises principally from the relationship between vegetation structure and aspect). For this exercise, we assumed that grazing patterns were the same in 1982 as in 2000.

We applied parameters to Hanski's incidence function model¹³ using the 2000 distribution of *H. comma* in Surrey (86 occupied patches; 30 unoccupied). In the model,

extinction probability declines with increasing patch area, and colonization probability increases with proximity to existing populations. The MCMC method (1,000 function evaluations in initiation, 4,000 function evaluations in estimation) was used for parameter estimation¹⁵. Set parameters were $A_0 = 0.02$ ha, and $B = 0.5$; estimated parameters were $x = 0.278$, $y = 7.261$, $e = 0.337$, $\alpha = 0.445$. A_0 , e and x scale the relationship between patch area and extinction, B and α determine migration rates, and y determines the relationship between migration and colonization probability¹⁵. We assumed no colonization from outside the network and no regional stochasticity¹⁵. These independent parameter estimates were used to run 100 simulations of 18 generations each for both the 1982 and 2000 habitat models in the South Downs. Eight 1982 network simulations went completely extinct; one had only one surviving population. These were ignored in calculations of 1982 network expansion distances (including them results in an approximately 9% reduction in mean distance colonized).

Aricia agestis

Aricia agestis in Cornwall (Ordnance Survey (OS) grid reference SW5740) has been associated with *Erodium* and *Geranium* species for over 100 generations and is greater than 100 km from *H. chamaecistus*. The north Wales population (SH7683) has an equally long association with *H. chamaecistus*, which receives greater than 99% of eggs (*Erodium* and *Geranium* are present in the region but rare in *H. chamaecistus* habitat)^{16,19}. The new Barnack population (TF0704) uses *H. chamaecistus* on limestone (where Geraniaceae are rare). The new Gibraltar Point population (TF5659) uses *E. cicutarium*, and perhaps other Geraniaceae, in sand dunes (*H. chamaecistus* is absent). Barnack and Gibraltar Point are separated by 72 km; other inter-population distances are even greater (less than 5% *A. agestis* individuals are expected to move over 1 km per generation^{18,19}). Twenty-four plants each of *E. cicutarium*, *G. molle*, *Geranium dissectum* and *H. chamaecistus* were sunk to ground level in pots in each population in August 2000. We counted eggs over the following month. Numbers of eggs on an individual plant were treated as single observations, considering the effects of host plant, population and host × population interaction in a two-way Schreier–Ray–Hare test (equivalent to two-way non-parametric analysis of variance). Significant interactions reveal differences between populations in host choice. This test is overly conservative because some plant individuals of each species might never have been found by searching females, increasing count variation within each plant species. Therefore, we also used χ^2 tests, treating each egg laid as an independent observation—females can choose where to lay each egg independently.

For common environment rearing, eggs were obtained from wild-caught females from May to June 2000. Larvae from both populations were reared on the same mix of host plants at 18/6-h light/dark cycle (24°C light phase; 20°C dark phase). Emerging adults were released in July 2000, into a roughly 0.25-ha field that contained no natural host plants, into which had been sunk 432 plants each of *H. chamaecistus* and *G. molle*. Eggs could not be identified to population, so Cornish butterflies ($n = 162$ females) were released 10 days before Welsh butterflies ($n = 155$ females). Eggs were counted (and removed) just before release of the Welsh butterflies, and again at the end of the experiment. The first count contained only eggs from the Cornish butterflies, but a few Cornish females still survived when Welsh butterflies were released, so the Welsh result underestimates their true bias towards *H. chamaecistus*.

Bush crickets

Bush cricket wing morphologies were recorded (100% accuracy for *M. roeselii*; 98.25% for *C. discolor*, on the basis of detailed morphological measurements of a sample of 286 *C. discolor* specimens) in the field, from August to October 2000. The sexes did not differ in morphology frequencies. The year of the first record from each bush cricket population was obtained from refs 7–9 (also from J. Widgery, personal communication).

Received 18 January; accepted 6 March 2001.

1. Thomas, J. A. Holocene climate change and warm man-made refugia may explain why a sixth of British butterflies inhabit unnatural early-successional habitats. *Ecography* **16**, 278–284 (1993).
2. Thomas, J. A., Rose, R. J., Clarke, R. T., Thomas, C. D. & Webb, N. R. Intraspecific variation in habitat availability among ectothermic animals near their climatic limits and their centres of range. *Funct. Ecol.* **13** (Suppl.1), 55–64 (1999).
3. Hughes, L. Biological consequences of global warming: is the signal already apparent? *Trends Ecol. Evol.* **15**, 56–61 (2000).
4. Hill, J. K., Thomas, C. D. & Huntley, B. Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proc. R. Soc. Lond. B* **266**, 1197–1206 (1999).
5. Parmesan, C. *et al.* Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579–583 (1999).
6. Asher, J. *et al.* *The Millennium Atlas of Butterflies in Britain and Ireland: Provisional Maps* (Oxford Univ. Press, 2001).
7. Marshall, J. A. & Haes, E. C. M. *Grasshoppers and Allied Insects of Great Britain and Ireland* (Harley Books, Colchester, 1988).
8. Haes, E. C. M. & Harding, P. T. *Atlas of the Grasshoppers, Crickets and Allied Insects in Britain and Ireland* (The Stationery Office, London, 1997).
9. Widgery, J. *Orthoptera Recording Scheme for Britain and Ireland. Newsletter* 26 (BRC, Huntingdon, 2000).
10. Thomas, C. D. & Lennon, J. J. Birds extend their ranges northwards. *Nature* **399**, 213 (1999).
11. Cwynar, L. C. & MacDonald, G. M. Geographical variation of lodgepole pine in relation to population history. *Am. Nat.* **129**, 463–469 (1987).
12. Thomas, J. A., Thomas, C. D., Simcox, D. J. & Clarke, R. T. The ecology and declining status of the silver-spotted skipper butterfly (*Hesperia comma*) in Britain. *J. Appl. Ecol.* **23**, 365–380 (1986).
13. Hanski, I. A practical model of metapopulation dynamics. *J. Anim. Ecol.* **63**, 151–162 (1994).
14. Wahlberg, N., Moilanen, A. & Hanski, I. Predicting the occurrence of endangered species in fragmented landscapes. *Science* **273**, 1536–1538 (1996).
15. Moilanen, A. The equilibrium assumption in estimating the parameters of metapopulation models. *J. Anim. Ecol.* **69**, 143–153 (2000).
16. Heath, J., Pollard, E. & Thomas, J. A. *Atlas of Butterflies in Britain and Ireland* (Viking, Harmondsworth, 1984).
17. Perring, F. H. & Walters, S. M. *Atlas of the British Flora* 2nd edn (EP Publishing, Wakefield, 1976).
18. Bourn, N. A. D. & Thomas, J. A. The ecology and conservation of the brown argus butterfly *Aricia agestis* in Britain. *Biol. Conserv.* **63**, 67–74 (1993).
19. Wilson, R. J. *The Spatiotemporal Dynamics of Three Lepidopteran Herbivores of Helianthemum chamaecistus*. Thesis, Univ. Leeds (1999).
20. Thomas, C. D., Thomas, J. A. & Warren, M. S. Distributions of occupied and vacant butterfly habitats in fragmented landscapes. *Oecologia* **92**, 563–567 (1992).
21. Thomas, C. D. & Hanski, I. in *Metapopulation Biology: Ecology, Genetics and Evolution* (eds Hanski, I. A. & Gilpin, M. E.) 359–386 (Academic, London, 1997).
22. Niemela, J. & Spence, J. R. Distribution and abundance of an exotic ground-beetle (Carabidae): a test of community impact. *Oikos* **62**, 351–359 (1991).
23. Hill, J. K., Thomas, C. D. & Blakeley, D. S. Evolution of flight morphology in a butterfly that has recently expanded its geographic range. *Oecologia* **121**, 165–170 (1999).
24. Hill, J. K., Thomas, C. D. & Lewis, O. T. Flight morphology in fragmented populations of a rare British butterfly, *Hesperia comma*. *Biol. Conserv.* **87**, 277–284 (1999).
25. Zera, A. J. & Denno, R. F. Physiology and ecology of dispersal polymorphisms in insects. *Annu. Rev. Ent.* **42**, 207–230 (1997).
26. Doebeli, M. & Ruxton, R. D. Evolution of dispersal rates in metapopulation models: branching and cyclic dynamics in phenotype space. *Evolution* **51**, 1730–1741 (1997).
27. Coope, G. R. in *Extinction Rates* (eds Lawton, J. & May, R.) 55–74 (Oxford Univ. Press, 1995).
28. Williamson, M. *Biological Invasions* (Chapman & Hall, London, 1996).
29. Ratnieks, F. L. W. in *The "African" Honey Bee* (eds Spivak, M., Fletcher, D. J. C. & Breed, M. D.) 119–135 (Westview, Boulder, Colorado, 1991).
30. Thomas, C. D. & Jones, T. M. Partial recovery of a skipper butterfly (*Hesperia comma*) from population refuges: lessons for conservation in a fragmented landscape. *J. Anim. Ecol.* **62**, 472–481 (1993).

Acknowledgements

We thank D. Blakeley, H. Burton, C. Dytham, K. Ericson, P. Ewin, R. Fox, C. Gardener, S. Glencross, A. Goodhand, S. Hanna, I. Hanski, S. Hartley, J. Hill, C. Holloway, P. Hoskin, W. Kunin, R. Leaper, G. Jeffcoate, J. Lennon, J. Mellings, A. Moilanen, C. Parmesan, G. Parris, M. Rosenzweig, D. Sazer, J. Silvertown, M. Singer, D. Taneyhill, J. Thomas, J. Travis, J. Widgery, K. Wilson and I. Wynne for discussions, assistance and/or comments on the manuscript. Maps were produced in DMap. This work was funded by The UK Natural Environment Research Council, EU TMR FRAGLAND and English Nature. We also thank English Nature, the National Trust and other landowners for permission to carry out the research, and the amateur butterfly and Orthoptera recorders who recorded 10-km distributions.

Correspondence and requests for materials should be addressed to C.D.T. (e-mail: c.d.thomas@leeds.ac.uk).

Honeybee dances communicate distances measured by optic flow

Harald E. Esch^{*}, Shaowu Zhang[†], Mandyan V. Srinivasan[†] & Juergen Tautz[‡]

^{*} Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA

[†] Centre for Visual Science, Research School of Biological Sciences, Australian National University, Canberra, ACT 2601, Australia

[‡] Biozentrum Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany

In honeybees, employed foragers recruit unemployed hive mates to food sources by dances from which a human observer can read the distance and direction of the food source¹. When foragers collect food in a short, narrow tunnel, they dance as if the food source were much farther away. Dancers gauge distance by retinal image flow on the way to their destination. Their visually driven odometer misreads distance because the close tunnel walls increase optic flow². We examined how hive mates interpret these dances. Here we show that recruited bees search outside in the direction of the tunnel at exaggerated distances and not inside the tunnel where the foragers come from. Thus, dances must convey information about the direction of the food source and the total amount of image motion *en route* to the food source, but they do not convey information about absolute distances. We also found that perceived distances on various outdoor routes from the same hive could be considerably different. Navigational errors are avoided as recruits and dancers tend to fly in the same direction. Reported racial differences in honeybee dances¹ could have arisen merely from differences in the environments in which these bees flew.

We began by setting up an 8-m tunnel pointing southwards, with its entrance 3 m from the hive. A fresh set of ten marked bees was trained to forage from a feeder placed at the far end of the tunnel (feeder distance 11 m). The mean waggle duration of tunnel dancers was 358 ms (Table 1). We then determined which distance corresponds to a waggle duration of 358 ms in bees that fly outdoors, outside the tunnel, to a feeder in the southern direction. Ten marked foragers were trained to fly to a feeder positioned successively at various distances due south, up to a maximum of 450 m. Several hundred dances of marked individuals returning from the feeder were videotaped at each position. The calibration curve relating waggle duration to feeder distance is shown in Fig. 1 (details in Table 1). According to this calibration, tunnel bees that waggle for 358 ms indicate an outdoor feeder 72 m south of the hive.

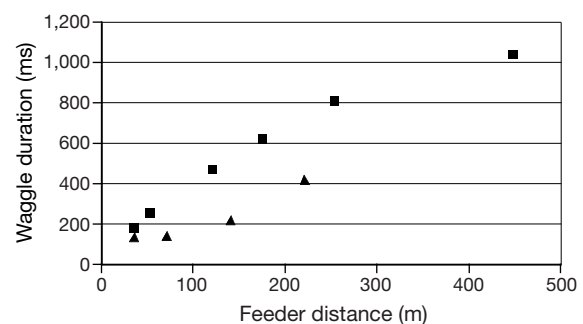


Figure 1 Distance calibration. Variation of waggle duration with feeder distance for an outdoor feeder positioned at various distances from the hive in the southern (squares) and northwestern direction (triangles).

Next, we examined how recruits interpret tunnel dances. Control stations were set up at 35, 70 and 140 m in the southern direction along the axis of the tunnel. In the first 30 min of each experiment, tunnel bees were prevented from entering the tunnel by closing its entrance. No dances occurred, and no recruits were observed at control stations.

Tunnel bees began to dance when the tunnel was opened. Newcomers appeared near control sites about 10 min thereafter. Observers at each control station counted the number of individuals searching persistently within a radius of 1 m from the control feeder. Recruits searched for a short time and disappeared. Their arrivals were spread over the whole observation period. We did not use scent, and newcomers were not rewarded in any way (see Methods); thus they had no obvious reason to stay around. If they returned, their persistence showed that they must have had information about feeder distance. Bees that alighted on control station feeders were caught and put into alcohol. This ensured that a tunnel bee had recruited each visiting bee and not a forager that had found a control station previously. Control sites were monitored for 2 h 30 min. The search distribution is shown in Fig. 2a. Most recruits appeared at the 70-m site. No recruits were observed at the tunnel feeder. Recruits must have read tunnel bee dances and interpreted them to represent a distance of 70 m—much greater than the 11 m the tunnel bees had actually flown (3 m outdoors and 8 m in the tunnel). Seventy metres is very close to the distance a human observer reads from tunnel dances (72 m). Thus, it would seem that most recruits read the objective distance information given by dancing tunnel bees and responded accordingly.

We repeated the above series of experiments with a 6-m-long tunnel pointing in the northwestern direction. In this case, bees that returned from the tunnel showed a mean waggle duration of 270 ms (Table 1). This duration is predicted by the results of the southerly tunnel, in which an 8-m flight elicited a mean waggle duration of 358 ms. That is, the waggle duration is proportional to the distance flown in the tunnel. From the calibration curve for the southern direction (Fig. 1), a waggle duration of 270 ms translates to an outdoor flight of 52 m.

To check whether recruited bees searched at the predicted distance, control sites were set up in the northwestern direction, as in the southern experiment. When the tunnel was closed for 30 min, no recruits appeared at the control sites. The tunnel was then opened and the control sites were monitored for 2 h 50 min. Surprisingly, most of the recruits searched at 140 m (Fig. 2b). In a second experiment, an additional control site was set up at 220 m, and recruits were monitored for 1 h 30 min. Most of the recruits again searched near 140 m (Fig. 2c).

Why did the recruited bees search at twice the distance we expected from the distance calibration curve of Fig. 1? To explore this, we measured a new distance calibration curve for the northwestern direction. In this case, the waggle duration increased much more slowly with distance, compared with the south-directed

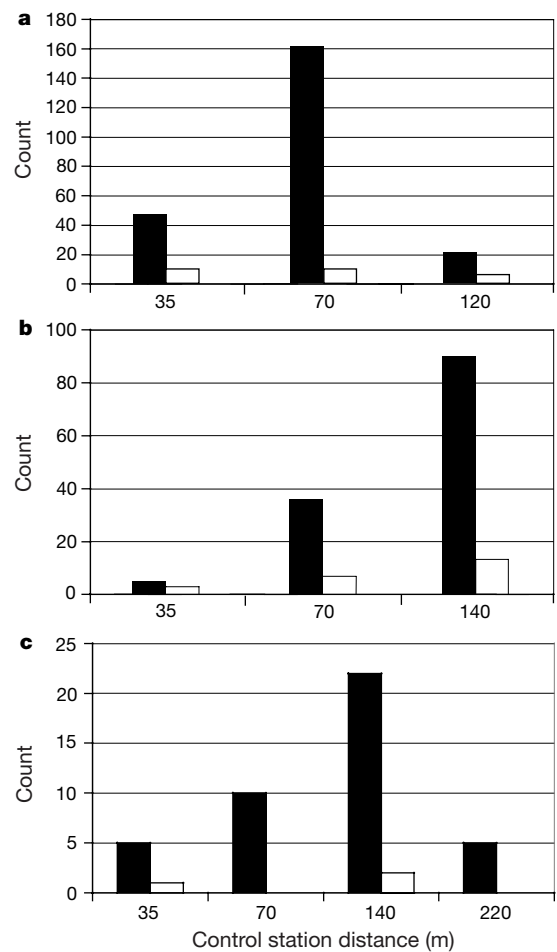


Figure 2 Distribution of counts of searching bees recruited by tunnel bees. Searching bees, black bars; bees caught, white bars. **a**, Control stations at three different distances from the hive in the southern direction. The tunnel pointed due south. **b**, Control stations at three different distances from the hive in the northwestern direction. The tunnel pointed due northwest. **c**, As in **b**, but with an additional station at 220 m.

calibration curve (Fig. 1; details in Table 1). From this calibration curve we see that a waggle duration of 270 ms corresponds to an outdoor flight distance of 157 m. This is in good agreement with the preferred control-site distance of 140 m (see above). Thus, in this series of experiments, the recruits also read the dances of tunnel bees like a human observer and responded accurately.

We conclude from our study that tunnel bees communicate the optic flow they perceived on the way to the feeder. Other hive mates

Table 1 Data for the duration of waggle runs

Experiment	Feeder distance (m)	Waggle duration (ms) (mean $\pm$ s.d.)	Number of bees analysed	Number of dances analysed	Number of waggle runs analysed
Calibration south	35	186 $\pm$ 75.3	3	23	214
	52	256 $\pm$ 73.3	6	30	434
	120	475 $\pm$ 123.4	17	89	992
	175	624 $\pm$ 111.4	13	43	418
	254	811 $\pm$ 116.8	4	10	142
	448	1,045 $\pm$ 127.1	3	4	127
Calibration northwest	35	143 $\pm$ 63.8	7	15	203
	70	152 $\pm$ 52.4	4	6	92
	140	229 $\pm$ 44	3	10	185
	220	424 $\pm$ 121.2	3	10	166
Tunnel south	11	358 $\pm$ 104.4	5	11	124
Tunnel northwest	9	270 $\pm$ 81.5	19	62	828

s.d. is the standard deviation of waggle run duration determined for different individuals.

can take this information and fly to locations that are reached with a similar optic flow. The fact that calibration curves are different for flights in the southern and northwestern directions implies that the distance calibration of a bee's 'odometer' is not absolute; rather, it depends on flight altitude and the nature of the landscape through which the bee flies. This is to be expected from a visually driven odometer^{2,3}. The flight path to control stations in the southern direction led up a slowly ascending hill. The 120-m control site was 10 m higher than the hive, and the 450-m control site was 30 m higher still. The route ran along a small paved road with some trees and wild flowers up to 1 m high on the left side of the road. Observers frequently saw the recruits approach along the paved road at an altitude of 1–2 m. In contrast, the trail in the northwestern direction led downhill over a meadow. The 140-m control site was 10 m lower than the hive. From 140 m on, the route ran along a fairly level paved road with high trees and bushes on both sides. Heran⁴ already showed that bees indicate an uphill feeder as farther away than a downhill feeder. He attributed this difference to different energy requirements on the way to the feeder ('the energy hypothesis') and not to differences in optic flow³.

The navigational information that a scout bee provides to a potential recruit in her waggle dance is fairly simple. The dance conveys only the direction of flight and the total amount of image motion that is expected to occur on the way. There does not seem to be any information on the nature of the environment in which the recruit must move (if there were, the recruits in our experiments would have ended up in the tunnel). In nature, therefore, the accuracy of distance information gained from dances depends critically on the directional information received at the same time, for it is the direction of flight that determines the environment through which a recruit flies. Clearly, then, there must be a high selection pressure to ensure that a dance signals the direction of the food source as precisely as possible. As long as the recruit flies in the same direction as the dancer, she will translate the image motion signalled by the dancer into the correct flight distance and find the goal.

Considering the large differences in dance behaviour that we have observed in the same hive in this study, future experiments ought to explore whether the reported racial differences in honeybee waggle dances¹ arose merely from differences in the environments in which the bees flew. □

Methods

Bees (*Apis mellifera ligustica*) were kept in a two-frame observation hive. All dances of returning foragers were recorded on videotape. Essential features of the tunnel setup were described elsewhere², but some dimensions were different. The tunnel was 6 or 8 m long, 8 cm wide and 18 cm high. The walls and floor of the tunnel were covered with a texture composed of random black and white 1 cm² pixels printed on paper. The top of the tunnel was covered with black insect screen mesh, and a feeder was placed at the far end. The feeder provided 2 M sugar solution in a glass vessel 2 cm high and 5 cm in diameter, placed inverted on a grooved plastic plate, 7 cm in diameter, on top of a yellow star 10 cm in diameter, painted on a blue background. No scent was used at the feeder. Foragers in the tunnel could see the sky through the mesh. Bees could enter and leave the tunnel only through the near end, as the far end was closed off. The tunnel and observation hive were outdoors, in hilly, open farmland at the Giesberg farm of the University of Würzburg.

The dances of foragers returning from the tunnel were videotaped and analysed frame-by-frame. Recordings were made with a digital camera and replayed at 25 frames per s. Waggle duration was determined by counting the number of frames per waggle run.

Identical settings were arranged at all control stations. The same type of feeder as used in the tunnel was filled with water and placed on a small wooden table (15 cm², 1 m above ground) on top of a yellow star painted on a blue background. A bright yellow camping chair next to the feeder made each control site conspicuous from a distance.

Received 23 February; accepted 3 April 2001.

1. Frisch, K. von. *The Dance Language and Orientation of Bees* (Harvard Univ. Press, Cambridge, Massachusetts, 1993).
2. Srinivasan, M. V., Zhang, S. W., Altwein, M. & Tautz, J. Honeybee navigation: nature and calibration of the "odometer". *Science* **287**, 851–853 (2000).
3. Esch, H. E. & Burns, J. E. Distance estimation by foraging honeybees. *J. Exp. Biol.* **99**, 155–162 (1996).

4. Heran, H. Ein Beitrag zur Frage nach der Wahrnehmungsgrundlage der Entfernungsweisung der Bienen. *Z. Vergl. Physiol.* **42**, 103–163 (1956).

Acknowledgements

This research was supported by grants from the Whitehall Foundation to H.E.E., from the Deutsche Forschungsgemeinschaft to J.T. and partly by grants from the Human Frontiers in Science Program, the US Defense Advanced Research Projects Agency and the Office of Naval research to M.V.S. We thank C. Baum, A. Meikus, A.-K. Scheuerman and C. Sylvester for their help during experiments and data evaluation.

Correspondence and requests for materials should be addressed to H.E. (e-mail: harald.e.esch.1@nd.edu).

Single cocaine exposure *in vivo* induces long-term potentiation in dopamine neurons

Mark A. Ungless*, Jennifer L. Whistler*, Robert C. Malenka† & Antonello Bonci*

* Ernest Gallo Clinic and Research Center, Department of Neurology, University of California, San Francisco, California 94110, USA

† Nancy Pritzker Laboratory, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, California 94134, USA

How do drugs of abuse modify neural circuitry and thereby lead to addictive behaviour? As for many forms of experience-dependent plasticity, modifications in glutamatergic synaptic transmission have been suggested to be particularly important^{1–4}. Evidence of such changes in response to *in vivo* administration of drugs of abuse is lacking, however. Here we show that a single *in vivo* exposure to cocaine induces long-term potentiation of AMPA (α -amino-3-hydroxy-5-methyl-isoxazole propionic acid)-receptor-mediated currents at excitatory synapses onto dopamine cells in the ventral tegmental area. Potentiation is still observed 5 but not 10 days after cocaine exposure and is blocked when an NMDA (N-methyl-D-aspartate) receptor antagonist is administered with cocaine. Furthermore, long-term potentiation at these synapses is occluded and long-term depression is enhanced by *in vivo* cocaine exposure. These results show that a prominent form of synaptic plasticity can be elicited by a single *in vivo* exposure to cocaine and therefore may be involved in the early stages of the development of drug addiction.

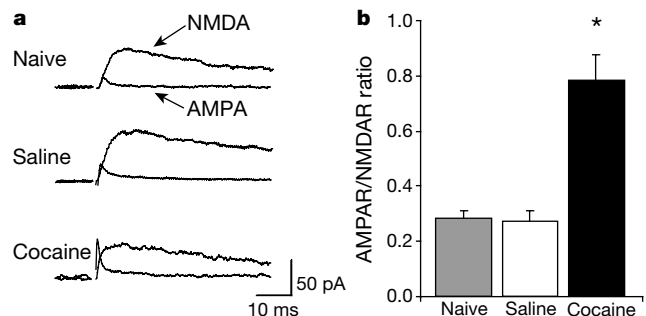


Figure 1 A single exposure to cocaine induced an increase in the AMPAR/NMDAR ratio of glutamatergic synaptic currents in VTA dopamine cells. **a**, Sample EPSCs in neurons from naive animals or animals pretreated with cocaine or saline. **b**, Peak AMPAR- and NMDAR-mediated EPSCs expressed as a ratio. Single cocaine injection ($n = 6$) results in a significantly increased AMPAR/NMDAR ratio compared with saline-injected ($n = 6$; $P < 0.05$) or naive animals ($n = 5$; $P < 0.05$).

The midbrain dopamine neurons and their widespread projections to areas such as the nucleus accumbens, prefrontal cortex and amygdala are involved in both acute and chronic responses to drugs of abuse^{1–4}. Glutamatergic systems are important in these responses; in particular, potentiation of glutamatergic synaptic transmission in the ventral tegmental area (VTA) may be involved in behavioural sensitization^{2,4}, a phenomenon thought to be important for the development of addiction. Long-term potentiation (LTP) and long-term depression (LTD) of synaptic transmission have been observed *in vitro* in the VTA^{5–8}, but synaptic plasticity as a consequence of *in vivo* drug administration has not been demonstrated. We directly examined these synapses in mice following a single exposure to cocaine.

Stimulation of excitatory afferents in the VTA activates both postsynaptic AMPA receptors (AMPA) and NMDA receptors (NMDARs)^{5,9}. We compared the relative contribution of AMPARs and NMDARs to excitatory postsynaptic currents (EPSCs) recorded in dopamine cells in midbrain slices from naive mice, mice injected with saline solution the previous day and mice injected with cocaine (15 mg kg⁻¹) the previous day. EPSCs were evoked while holding neurons in voltage-clamp at +40 mV in the absence and then the presence of the NMDAR antagonist AP5 (D(-)-2-amino-5-phosphonovaleric acid; 50 μ M), and an AMPAR/NMDAR ratio was computed (see Methods; Fig. 1a). Mice injected with a single dose of cocaine exhibited a significantly larger AMPAR/NMDAR ratio

(0.77 ± 0.11 ; $n = 6$) than did mice injected with saline (0.27 ± 0.04 ; $n = 6$; $P < 0.05$) or naive mice (0.28 ± 0.03 ; $n = 5$; $P < 0.05$), which did not differ significantly from one another ($P > 0.05$; Fig. 1b).

A change in the relative contribution of AMPARs and NMDARs to EPSCs probably reflects an increase in AMPAR function or number, or a decrease in NMDAR function or number, or a combination of these¹⁰. However, such changes do not rule out the possibility that additional modifications in transmitter release mechanisms may be induced by cocaine. To determine whether changes in the probability of transmitter release (P_r) occurred in the cocaine-treated animals, we compared the responses to paired pulses, a measure that changes in a highly predictable fashion with changes in P_r ^{10,11}. The paired-pulse ratio did not differ between cocaine- and saline-treated mice when tested at a variety of intervals, suggesting that the *in vivo* cocaine exposure did not cause a change in P_r (Fig. 2a).

To investigate whether AMPAR function and/or number was modified in the cocaine-treated mice, we examined miniature AMPAR-mediated EPSCs (mEPSCs), a standard method for determining the locus of synaptic change¹². Cocaine-treated mice showed a significant increase in both amplitude and frequency (17.4 ± 2.4 pA; 2.4 ± 0.7 Hz; $n = 12$) of mEPSCs, compared with saline controls (13.9 ± 0.7 pA; $P < 0.05$; 0.7 ± 0.2 Hz; $P < 0.05$ (Mann–Whitney); $n = 7$; Fig. 2b–f). An increase in the amplitude of mEPSCs is routinely attributed to a postsynaptic change in receptor function, number or both. An increase in the frequency of mEPSCs could be due to either an increase in P_r that was not detected using the paired-pulse analysis or a postsynaptic increase in the number of functional AMPARs at synapses that were previously electrophysiologically ‘silent’¹³. Furthermore, under certain circumstances increases in mEPSC amplitude and frequency can be caused by increases in the concentration of glutamate in the synaptic cleft^{14,15}. Therefore, we tested whether *in vivo* cocaine administration induced postsynaptic increases in AMPAR function or number by exogenously applying AMPA.

When AMPA (1 μ M for 30 s) was superfused onto neurons from

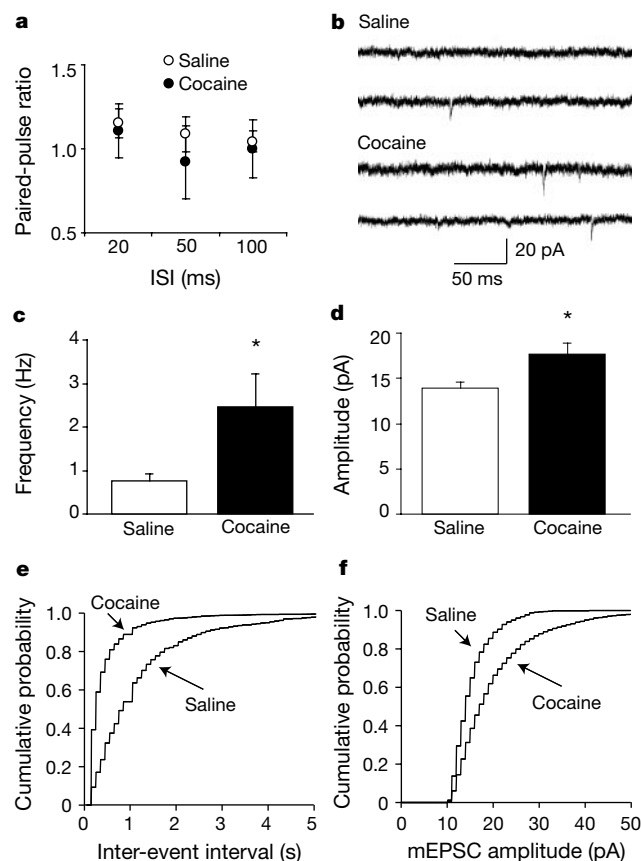


Figure 2 Cocaine exposure had no effect on paired-pulse modulation but increased mEPSC frequency and mEPSC amplitude. **a**, Paired-pulse protocols showed no differences between groups ($n = 5$ and 4) at a variety of inter-stimulus intervals (ISI; $P > 0.05$). **b**, Samples of mEPSCs in neurons from animals pretreated with saline or cocaine. **c**, Cocaine ($n = 12$) induced a significant increase in mEPSC frequency compared with saline ($n = 7$; $P < 0.05$). **d**, mEPSCs were also larger in cocaine-treated animals ($n = 12$) compared with saline-treated animals ($n = 7$; $P < 0.05$). **e**, **f**, Cumulative probability plots of amplitude and frequency for all mEPSCs in cocaine- and saline-treated animals.

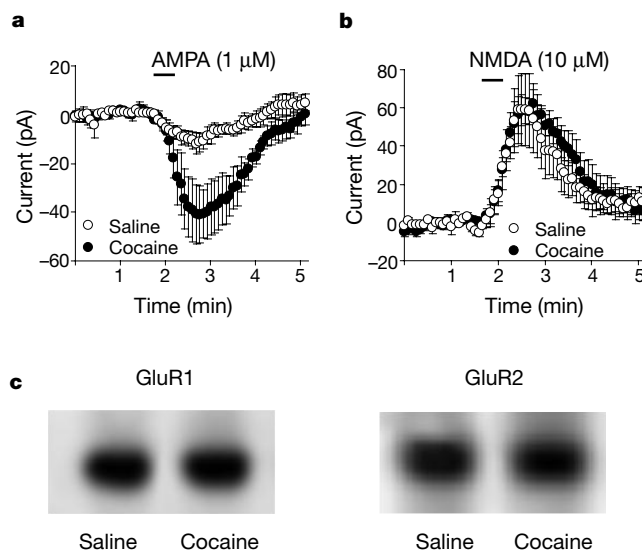


Figure 3 Cocaine exposure caused a larger response to AMPA but not NMDA and did not change the total amount of AMPAR subunits. **a**, Bath application of AMPA (1 μ M) for 30 s elicited greater inward current (5-s bins) in cocaine-treated animals ($n = 5$) than in saline-treated animals ($n = 5$; $P < 0.05$). **b**, Bath application of NMDA (10 μ M) for 30 s elicited similar current (5-s bins) in both groups of animals ($n = 5$ and 4 ; $P > 0.05$) when neurons were voltage-clamped at +40 mV. **c**, VTA slices taken from cocaine- and saline-treated animals exhibited similar expression of GluR1 and GluR2 receptors. Immunoblots were quantified by scanning densitometry. Signal from cocaine- and saline-treated animals differed by less than 5% for both GluR1 and GluR2. Immunoblots are representative of three independent experiments.

the cocaine-treated mice the AMPA-induced currents were significantly increased (-40 ± 12 pA; 2.5–3.0 min; $n = 5$) over those induced in mice injected with saline (-11 ± 3 pA; $n = 5$; $P < 0.05$; Fig. 3a). This difference provides strong evidence that a postsynaptic change in AMPAR number, function or both had occurred. To determine whether, in addition, cocaine induced a significant decrease in NMDAR function or number, we repeated similar experiments using bath application of NMDA ($10 \mu\text{M}$ for 30 s while holding cells at $+40$ mV). In contrast to the effects of AMPA, there was no difference between the NMDA-induced currents in cocaine-injected mice (58 ± 6 pA; $n = 5$) and those from the mice administered saline (54 ± 18 pA; $n = 4$; $P > 0.05$; Fig. 3b).

These results indicate that the number, function or both of AMPARs in the postsynaptic membrane are increased following cocaine exposure. A single injection of cocaine does not change the total expression of AMPAR subunits¹⁶. Consistent with this, we found no change in GluR1 or GluR2 immunoreactivity, determined by western blotting, in VTA slices from mice injected with a single cocaine dose compared with mice injected with saline (Fig. 3c). This result suggests that cocaine induces either post-translational redistribution of existing AMPARs or phosphorylation of membrane AMPARs, similar to that seen during LTP in hippocampal neurons^{17–19}.

Although the VTA is considered to be essential for the development of behavioural sensitization to psychostimulants, its involvement is thought to be transient^{2–4,20,21}. Therefore, we examined how long the cocaine-induced change in the AMPAR/NMDAR ratio lasted. Slices from mice given a single injection of cocaine or saline were prepared 5 or 10 days after the injections. There remained a significant difference between the AMPAR/NMDAR ratios 5 days (cocaine 0.55 ± 0.05 ; saline 0.18 ± 0.07 ; $n = 5$ and 5; $P < 0.05$; Fig. 4a) but not 10 days after injection (cocaine 0.28 ± 0.05 ; saline 0.27 ± 0.05 ; $n = 5$ and 5; $P > 0.05$; Fig. 4a). This indicates that the cocaine-induced synaptic potentiation is transient, similar to other changes observed in the VTA following repeated drug exposures^{20,21}.

We next examined the regional specificity of cocaine's action by measuring AMPAR/NMDAR ratios in CA1 pyramidal cells in the hippocampus, an area that exhibits robust LTP *in vitro*. As shown in Fig. 4b, no significant change in the ratio was detected in cocaine-treated mice (1.13 ± 0.39 ; $n = 5$) compared to the saline-treated group (1.23 ± 0.59 ; $n = 4$; $P > 0.05$; Fig. 4b). As an additional test of the specificity of cocaine's actions, we measured AMPAR/NMDAR ratios in GABA (γ -aminobutyric acid) neurons in the VTA, which do not exhibit LTP *in vitro*⁵. GABA neurons from mice injected with a single dose of cocaine exhibited similar AMPAR/NMDAR ratios (0.31 ± 0.1 ; $n = 4$) to those of GABA neurons from mice injected with saline (0.29 ± 0.04 ; $n = 3$; $P > 0.05$; Fig. 4c). Further evidence

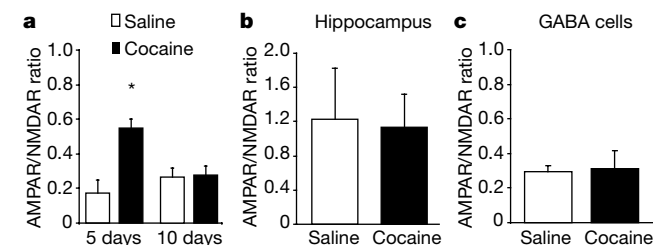


Figure 4 Cocaine-induced increase in AMPAR/NMDAR ratio in the VTA lasted for 5 but not 10 days after injection and did not occur in the hippocampus or in GABA neurons in the VTA. **a**, AMPAR/NMDAR ratios were still potentiated 5 days after a single injection of cocaine ($n = 5$) compared with saline ($n = 5$; $P < 0.05$). No differences were observed 10 days after injection of cocaine ($n = 5$) or saline ($n = 5$; $P > 0.05$). **b**, AMPAR/NMDAR ratios were similar in CA1 hippocampal neurons from animals receiving cocaine ($n = 5$) or saline ($n = 4$; $P > 0.05$). **c**, AMPAR/NMDAR ratios were similar in GABA neurons from animals receiving cocaine ($n = 4$) or saline ($n = 3$; $P > 0.05$).

that this cocaine-induced potentiation is region-specific comes from the observation that in the nucleus accumbens cocaine causes a decrease, not an increase, in the AMPAR/NMDAR ratio²².

A single exposure to cocaine can induce context-dependent behavioural sensitization^{23,24}, as we have confirmed in our mice. Mice injected with cocaine (15 mg kg^{-1}) and then given a challenge injection of cocaine (15 mg kg^{-1}) the following day exhibited significantly more locomotor activity ($9,343 \pm 1,694$ cm in 15-min trial; $n = 9$) than mice that had received saline the previous day and were challenged with cocaine ($4,661 \pm 1,219$ cm in 15-min trial; $n = 9$; $P < 0.05$). This effect is blocked by infusions of the NMDAR antagonist MK-801 directly into the VTA²⁴ (or when MK-801 is given systemically). Therefore, if the cocaine-induced increase in synaptic strength in the VTA is related to behavioural sensitization, administration of MK-801 with cocaine should block the change in the AMPAR/NMDAR ratio. Figure 5a shows that administration of MK-801 with cocaine did block the cocaine-induced change in the AMPAR/NMDAR ratio (0.27 ± 0.05 ; $n = 5$; Fig. 5a), and that MK-801 itself had no effect on the AMPAR/NMDAR ratio when injected together with saline (0.35 ± 0.06 ; $n = 8$; $P > 0.05$; Fig. 5a).

The induction of LTP *in vitro* in the VTA also depends upon NMDAR activation^{5,6}, and NMDAR-dependent LTP is thought to involve increases in AMPAR-function and number^{12,13,17–19}. These results indicate that LTP and cocaine-induced potentiation share similar mechanisms of expression and induction. If *in vivo* cocaine pretreatment induced LTP, then the *in vitro* induction of LTP should be reduced in slices prepared from these animals. Consistent with this prediction, we found that LTP was absent in cocaine-treated animals ($107 \pm 8\%$; 45–50 min; $n = 8$) compared with LTP in saline controls ($130 \pm 7\%$; $n = 6$; $P < 0.05$; Fig. 5b, d). If *in vivo* cocaine treatment has increased synaptic strength at these synapses, as our results indicate, then we might expect LTD to be enhanced, as has been shown in motor cortex following motor skill learning²⁵. We therefore examined LTD by administering a series of three LTD induction protocols (separated by 5 min) in an attempt to induce

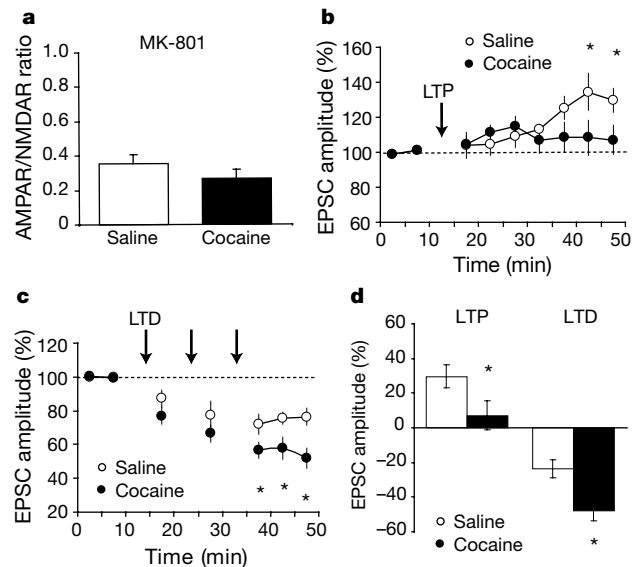


Figure 5 Cocaine-induced potentiation was blocked by the NMDAR antagonist MK-801, occluded LTP and enhanced LTD. **a**, Animals receiving MK-801 exhibited similar AMPAR/NMDAR ratios when co-injected with either cocaine ($n = 5$) or saline ($n = 8$; $P > 0.05$). **b**, Time course of mean EPSCs before and after LTP pairing protocol. LTP was significantly reduced in cocaine-treated animals ($n = 8$) compared with saline-treated controls ($n = 6$; $P < 0.05$). **c**, Time course of mean EPSCs before and after three LTD pairing protocols. LTD was significantly enhanced in cocaine-treated animals ($n = 5$) compared with saline controls ($n = 6$; $P < 0.05$). **d**, Summary of the magnitude of LTP and LTD in saline- and cocaine-treated animals at 45–50 min.

maximal, saturated LTD. Using this procedure we observed greater LTD in the cocaine-treated animals ($52 \pm 6\%$; $45\text{--}50\text{ min}$; $n = 5$) than in saline controls ($76 \pm 5\%$; $n = 6$; $P < 0.05$; Fig. 5c, d).

These experiments provide direct evidence that excitatory synapses in the VTA are potentiated following a single *in vivo* exposure to cocaine and that these changes closely resemble LTP. Because the neural mechanisms that underlie behavioural sensitization have primarily been studied in rat, we also examined the AMPAR/NMDAR ratio in Sprague–Dawley rats (P21–28) following a single *in vivo* injection of cocaine or saline. Consistent with our results in mice, the cocaine-treated animals exhibited significantly potentiated AMPAR/NMDAR ratios (saline 0.42 ± 0.05 ; cocaine 0.68 ± 0.1 ; $n = 10$ and 11 ; $P < 0.05$). The mechanisms by which cocaine potentiates excitatory synapses in the VTA remain to be determined. Two contributing factors may be the brief increase in glutamate release in the VTA that occurs immediately after cocaine exposure²⁶ and the dopamine-mediated inhibition of LTD^{7,8}. Consistent with our results, chronic *in vivo* cocaine administration increases the probability of VTA neurons reaching depolarization block in response to AMPA²¹. That synaptic plasticity at excitatory synapses on midbrain dopamine cells may generally contribute to the behavioural changes elicited by drugs of abuse is suggested by two additional findings. First, overexpression of the AMPAR subunit GluR1 in the VTA increases the stimulant and rewarding properties of morphine²⁷. Second, application of nicotine *in vitro* to slices of the VTA when paired with postsynaptic depolarization can elicit LTP²⁸. It will therefore be interesting to see whether morphine and nicotine when administered *in vivo* induce changes similar to those reported here.

The potentiation of excitatory inputs onto VTA neurons that we have observed would be expected to have substantial effects on activity within the dopamine system, and consequently on a wide range of behaviours related to drug abuse. In particular, this potentiation might represent a component of the neural basis of sensitization of the incentive-motivational system²⁹, a process thought to be central to the development of compulsive drug-seeking behaviour. In addition, the cocaine-induced potentiation may be important not just for the early stages of addiction, but also for relapse, where a single exposure to cocaine after a period of abstinence can induce renewed drug-seeking behaviour³⁰. Regardless of the behavioural consequences of LTP in the VTA, our results indicate that the cellular mechanisms thought to be involved in adaptive forms of experience-dependent plasticity may be maladaptively usurped by drugs of abuse. □

Methods

Electrophysiology

We prepared VTA and hippocampal slices from day 21–35 C57 mice (Charles River) as described^{5,7}. Mice were anaesthetized with halothane and then killed. A block of tissue containing midbrain was sliced in the horizontal plane ($230\text{ }\mu\text{m}$) in ice-cold low- Ca^{2+} solution (containing in mM: 126 NaCl, 1.6 KCl, 1.2 NaH_2PO_4 , 1.2 MgCl_2 , 0.625 CaCl_2 , 18 NaHCO_3 and 11 glucose). Slices were transferred to a holding chamber containing artificial cerebrospinal fluid (ACSF, in mM: 126 NaCl, 1.6 KCl, 1.2 NaH_2PO_4 , 1.2 MgCl_2 , 2.5 CaCl_2 , 18 NaHCO_3 and 11 glucose) and equilibrated at $31\text{--}34^\circ\text{C}$ for at least 1 h. Coronal hippocampal slices ($350\text{ }\mu\text{m}$) were made in ice-cold ACSF, and stored and recorded from at room temperature. Picrotoxin ($100\text{ }\mu\text{M}$) was added to the ACSF for recording, to block GABA_A-receptor-mediated inhibitory postsynaptic potentials. The mEPSCs were collected in the presence of lidocaine ($500\text{ }\mu\text{M}$), AP5 ($50\text{ }\mu\text{M}$) and sucrose (100 mM). All solutions were bubbled with $95\%\text{ O}_2/5\%\text{ CO}_2$, and perfused over the slice at 2.5 ml min^{-1} . Cells were visualized with an upright microscope using infrared illumination and whole-cell voltage-clamp recordings were made with an Axopatch 1D amplifier (Axon Instruments). Electrodes ($2\text{--}6\text{ M}\Omega$) contained in mM: 117 caesium methanesulphonic acid, 20 HEPES, 0.4 EGTA, 2.8 NaCl, 5 TEA-Cl, 2.5 MgATP , 0.25 MgGTP (pH $7.2\text{--}7.4$, $275\text{--}285\text{ mOsm}$). For perforated-patch LTP experiments, Amphotericin B (15 mg ml^{-1}) dissolved in dimethyl sulphoxide was added to internal solution (0.15% final concentration) containing in mM: 117 caesium methanesulphonic acid, 20 HEPES, 0.4 EGTA, 2.8 NaCl, 5 TEA-Cl (pH $7.2\text{--}7.4$, $275\text{--}285\text{ mOsm}$). Dopamine cells were identified by the presence of a large I_h current. GABA cells were identified by the absence of I_h . A bipolar stimulating electrode was placed $100\text{--}300\text{ }\mu\text{m}$ rostral to the recording electrode and was used to stimulate excitatory afferents at 0.1 Hz . Neurons were voltage-clamped at -70 mV , except where noted. LTP was elicited by holding cells at $+10\text{ mV}$ for 200 stimuli at 1 Hz .

LTD was elicited by holding cells at -40 mV for 400 stimuli at 1 Hz . Data were analysed in 5-min bins using a two-way repeated measures analysis of variance and Student–Newman–Keuls post hoc tests. We recorded from hippocampal CA1 neurons in a similar manner and EPSCs were evoked by stimulating the stratum radiatum. VTA slices from rats were prepared and recorded from as reported⁵, except that the internal pipette solution was the same as that used for mice.

EPSCs were filtered at 2 kHz , digitized at $5\text{--}10\text{ kHz}$ and stored using IgorPro software (WaveMetrics). To calculate the AMPAR/NMDAR ratio, an average was computed of EPSCs at $+40\text{ mV}$ (12 EPSCs) before and after application of AP5 ($50\text{ }\mu\text{M}$) for 5 min. The average response in the presence of AP5 (AMPA-only) was subtracted from that seen in its absence and an average NMDAR-only response was calculated. Then the peak of the AMPAR-only EPSC was divided by the peak of the NMDAR-only EPSC to give an AMPAR/NMDAR ratio. Miniature EPSCs were collected using Clampex (Axon Instruments) and analysed using Mini Analysis Program (Synaptosoft). Detection criteria were set at $>10\text{ pA}$ amplitude, $<1\text{ ms}$ rise-time and $<3\text{ ms}$ decay-time. All values are expressed as mean $\pm$ s.e.m. Statistical significance was assessed using two-tailed Student's *t*-tests, except where stated otherwise. In each set of experiments, there were no differences in the results from blinded and non-blinded experiments and therefore results were combined.

Immunoblot

Slices of VTA were prepared in ACSF exactly as for electrophysiology. Slices were disrupted in $100\text{ }\mu\text{l}$ ACSF in the presence of protease inhibitors on ice using a teflon/glass dounce. The dounce was washed with $100\text{ }\mu\text{l}$ ACSF with protease inhibitors and the wash was pooled with the homogenate. Samples were denatured in SDS sample buffer at 95°C for 2 min, separated by SDS–PAGE, transferred to nitrocellulose and immunoblotted with GluR1 (Chemicon PAB1504) or GluR2 (Chemicon MAB397) antibodies followed by incubation with HRP-conjugated secondary antibodies (Jackson ImmunoResearch) and development with ECL plus (Amersham).

Received 7 February; accepted 6 April 2001.

- Berke, J. D. & Hyman, S. E. Addiction, dopamine, and the molecular mechanisms of memory. *Neuron* **25**, 515–532 (2000).
- Wolf, M. E. The role of excitatory amino acids in behavioral sensitization to psychomotor stimulants. *Prog. Neurobiol.* **54**, 679–720 (1998).
- Kalivas, P. W. Interactions between dopamine and excitatory amino acids in behavioral sensitization to psychostimulants. *Drug Alcohol Depend.* **37**, 95–100 (1995).
- Clark, D. & Overton, P. G. Alterations in excitatory amino acid-mediated regulation of midbrain dopaminergic neurons induced by chronic psychostimulant administration and stress: relevance to behavioural sensitization and drug addiction. *Addict. Biol.* **3**, 109–135 (1998).
- Bonci, A. & Malenka, R. C. Properties and plasticity of excitatory synapses on dopaminergic and GABAergic cells in the ventral tegmental area. *J. Neurosci.* **19**, 3723–3730 (1999).
- Overton, P. G., Richards, C. D., Berry, M. S. & Clark, D. Long-term potentiation at excitatory amino acid synapses on midbrain dopamine neurons. *NeuroReport* **10**, 221–226 (1999).
- Thomas, M. J., Malenka, R. C. & Bonci, A. Modulation of long-term depression by dopamine in the mesolimbic system. *J. Neurosci.* **20**, 5581–5586 (2000).
- Jones, S., Kornblum, J. L. & Kauer, J. A. Amphetamine blocks long-term synaptic depression in the ventral tegmental area. *J. Neurosci.* **20**, 5575–5580 (2000).
- Johnson, S. W. & North, R. A. Two types of neurons in the rat ventral tegmental area and their synaptic inputs. *J. Physiol.* **450**, 455–468 (1992).
- Perkel, D. J. & Nicoll, R. A. Evidence for all-or-none regulation of neurotransmitter release: implications for long-term potentiation. *J. Physiol.* **471**, 481–500 (1993).
- Dobrunz, L. E. & Stevens, C. F. Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron* **18**, 995–1008 (1997).
- Malenka, R. C. & Nicoll, R. A. Long-term potentiation—a decade of progress? *Science* **285**, 1870–1874 (1999).
- Isaac, J. T., Nicoll, R. A. & Malenka, R. C. Evidence for silent synapses: implications for the expression of LTP. *Neuron* **15**, 427–434 (1995).
- Liu, G., Choi, S. & Tsien, R. W. Variability of neurotransmitter concentration and nonsaturation of postsynaptic AMPA receptors at synapses in hippocampal cultures and slices. *Neuron* **22**, 395–409 (1999).
- Choi, S., Klingauf, J. & Tsien, R. W. Postfusional regulation of cleft glutamate concentration during LTP at 'silent synapses'. *Nature Neurosci.* **3**, 330–336 (2000).
- Fitzgerald, L. W., Ortiz, J., Hamedani, A. G. & Nestler, E. J. Drugs of abuse and stress increase the expression of GluR1 and NMDAR1 glutamate receptor subunits in the rat ventral tegmental area: common adaptations among cross-sensitizing agents. *J. Neurosci.* **16**, 274–282 (1996).
- Shi, S. H. *et al.* Rapid spine delivery and redistribution of AMPA receptors after synaptic NMDA receptor activation. *Science* **284**, 1811–1816 (1999).
- Barria, A., Muller, D., Derkach, V., Griffith, L. C. & Soderling, T. R. Regulatory phosphorylation of AMPA-type glutamate receptors by CaM–KII during long-term potentiation. *Science* **276**, 2042–2045 (1997).
- Lee, H. K., Barbarosie, M., Kameyama, K., Bear, M. F. & Huganir, R. L. Regulation of distinct AMPA receptor phosphorylation sites during bidirectional synaptic plasticity. *Nature* **405**, 955–959 (2000).
- White, F. J. & Wang, R. Y. Electrophysiological evidence for A10 dopamine autoreceptor subsensitivity following chronic D-amphetamine treatment. *Brain Res.* **309**, 283–292 (1984).
- Zhang, X. F., Hu, X. T., White, F. J. & Wolf, M. E. Increased responsiveness of ventral tegmental area dopamine neurons to glutamate after repeated administration of cocaine or amphetamine is transient and selectively involves AMPA receptors. *J. Pharmacol. Exp. Ther.* **281**, 699–706 (1997).
- Thomas, M. J., & Malenka, R. C. Behavioral sensitization to cocaine is associated with changes in nucleus accumbens excitatory synaptic transmission. *Soc. Neurosci. Abstr.* **29.2.3** (2000).
- Jackson, H. C. & Nutt, D. J. A single preexposure produces sensitization to the locomotor effects of cocaine in mice. *Pharmacol. Biochem. Behav.* **45**, 733–735 (1993).
- Kalivas, P. W. & Alesdatter, J. E. Involvement of N-methyl-D-aspartate receptor stimulation in the ventral tegmental area and amygdala in behavioral sensitization to cocaine. *J. Pharmacol. Exp. Ther.* **267**, 486–495 (1993).

25. Rioult-Pedotti, M.-S., Friedman, D. & Donoghue, J. P. Learning-induced LTP in neocortex. *Science* **290**, 533–536 (2000).
26. Kalivas, P. W. & Duffy, P. D1 receptors modulate glutamate transmission in the ventral tegmental area. *J. Neurosci.* **15**, 5379–5388 (1995).
27. Carlezon, W. A. *et al.* Sensitization to morphine induced by viral-mediated gene transfer. *Science* **277**, 812–814 (1997).
28. Mansvelder, H. D. & McGehee, D. S. Long-term potentiation of excitatory inputs to brain reward areas by nicotine. *Neuron* **27**, 349–357 (2000).
29. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Brain Res. Rev.* **18**, 247–291 (1993).
30. de Wit, H. & Stewart, J. Reinstatement of cocaine-reinforced responding in the rat. *Psychopharmacology* **75**, 134–143 (1981).

Acknowledgements

We thank S. Nicola and H. Fields for comments on this manuscript, members of the Bonci and Malenka labs for discussions, and J. Kim for help with injections. This work was supported by funds provided by the State of California for medical research on alcohol and substance abuse through the University of California, San Francisco (A.B. and J.L.W.) and by grants from the NIH (R.C.M.).

Correspondence and requests for materials should be addressed to A.B. (e-mail: bonci@itsa.ucsf.edu).

Presynaptic glycine receptors enhance transmitter release at a mammalian central synapse

Rostislav Turecek & Laurence O. Trussell

Oregon Hearing Research Center and Vollum Institute, L-335A, 3181 SW Sam Jackson Park Rd., Portland, Oregon 97201, USA

Glycine and GABA_A (γ-aminobutyric acid A) receptors are inhibitory neurotransmitter-gated Cl[−] channels localized in post-synaptic membranes. In some cases, GABA_A receptors are also found presynaptically, but they retain their inhibitory effect as their activation reduces excitatory transmitter release^{1–4}. Here we report evidence for presynaptic ionotropic glycine receptors, using pre- and postsynaptic recordings of a calyceal synapse in the medial nucleus of the trapezoid body (MNTB). Unlike the classical action of glycine, presynaptic glycine receptors triggered a weakly depolarizing Cl[−] current in the nerve terminal. The depolarization enhanced transmitter release by activating Ca²⁺ channels and increasing resting intraterminal Ca²⁺ concentrations. Repetitive activation of glycinergic synapses on MNTB neurons also enhanced glutamatergic synaptic currents, indicating that presynaptic glycine receptors are activated by glycine spillover. These results reveal a novel site of action of the transmitter glycine, and indicate that under certain conditions presynaptic Cl[−] channels may increase transmitter release.

Each principal cell of the MNTB, an auditory brainstem nucleus, receives a single, large somatic excitatory synapse (the calyx of Held) whose activation evokes an excitatory postsynaptic current (EPSC) mediated primarily by AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors^{5,6}. When postsynaptic neurons were recorded using a CsF-based patch pipette solution and voltage clamped at −60 mV, postsynaptic responses to GABA or glycine were suppressed⁷, allowing us to monitor transmitter regulation of EPSC amplitude. Application of GABA reduces EPSCs because of the inhibition of transmitter release by presynaptic metabotropic GABA_B receptors^{8,9} (Fig. 1a). Glycine, however, produced the opposite effect, increasing the EPSC by about 40% (Fig. 1b, d; 1 mM glycine, 43 ± 15% increase, *n* = 20; 50 μM glycine, 35 ± 8% increase, *n* = 4; *P* < 0.01 by paired *t*-test for both concentrations). Enhancement of transmission by glycine quickly subsided upon

removal of glycine, was blocked by strychnine, an antagonist of glycine receptors, and could be evoked repeatedly with no significant decline for up to an hour of recording (Fig. 1b–d). The degree of potentiation by glycine was relatively constant in animals between postnatal days (P) 10 and P16 (coefficient of determination *r*² = 0.001 for relation between percentage of enhancement and age), making it unlikely that it subsides upon maturation of the auditory pathway.

Several observations indicated that glycine acted presynaptically and in a Ca²⁺-dependent manner. Paired-pulse ratios (EPSC₂/EPSC₁, 10 ms interval) were significantly reduced by glycine (Fig. 2a, b), indicating enhanced synaptic depression and that glycine may act presynaptically to increase the probability of transmitter release¹⁰. Moreover, application of glycine induced a significant increase in the frequency, but not in the amplitude, of spontaneous miniature EPSCs (mEPSCs) (Fig. 2c, d; *n* = 10, mean frequency in control, 1.7 ± 1.0 Hz; in glycine, 2.6 ± 1.5 Hz, significantly different at *P* < 0.01 with paired *t*-test; mean amplitude in control −69 ± 13 pA; in glycine −67 ± 14 pA, *P* > 0.45 by paired *t*-test). Increase in mEPSC frequency by glycine was prevented with Cd²⁺, a non-selective Ca²⁺-channel blocker (*n* = 4; Fig. 2d). In these experiments, Cd²⁺ alone did not change mEPSC frequency or amplitude (change in frequency, 5 ± 19%; in amplitude, 13 ± 16%, *P* > 0.2). Glycine's action on evoked EPSCs was examined in slices perfused for 10 min with 200 μM EGTA-AM, a membrane-permeant Ca²⁺ chelator^{11,12}. Postsynaptic recording pipettes contained 5 mM EGTA so that the effects on transmission observed with EGTA-AM were probably due to changes in presynaptic buffering. EGTA-AM decreased the amplitude of the EPSC by 29 ± 14%, indicating that Ca²⁺ ions had less access to Ca²⁺ sensors responsible for transmitter release¹³. More importantly, enhancement of Ca²⁺ buffering also eliminated the effect of glycine on release (Fig. 2e, f; EPSC_{glycine}/EPSC_{control} after EGTA-AM was 0.94 ± 0.07, *P* > 0.17, compared with 1.46 ± 0.18 before EGTA-AM treatment, *P* < 0.03, *n* = 4, paired *t*-test).

To determine the mechanism of glycine's action, patch-clamp recordings were made from the presynaptic terminal while glycine was applied. Glycine induced large strychnine-sensitive ionic currents whose reversal potential shifted with change in concentration of intracellular Cl[−] [Cl[−]]_i (from 17 mM to 150 mM, reversal

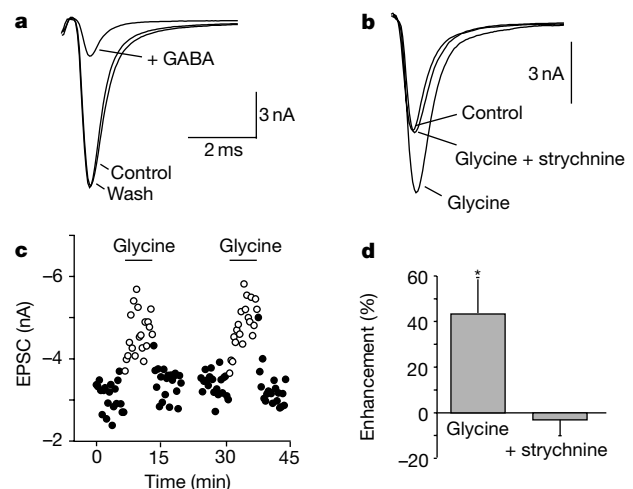


Figure 1 Enhancement of EPSCs by glycine. Application of 20 μM GABA inhibits EPSCs (a) whereas application of 1 mM glycine potentiates EPSCs, an effect blockable by 2 μM strychnine (b). Average of ten traces in a and b. c, Time course and magnitude of glycine enhancement in one experiment in which glycine was reapplied after 18 min. d, Mean effect of glycine (*n* = 20) and glycine plus strychnine (*n* = 6); asterisk, *P* < 0.001. Enhancement calculated as ((EPSC_{glycine}/EPSC_{control}) − 1) × 100. Cells loaded with CsF, held at −60 mV to minimize postsynaptic Cl[−] current.

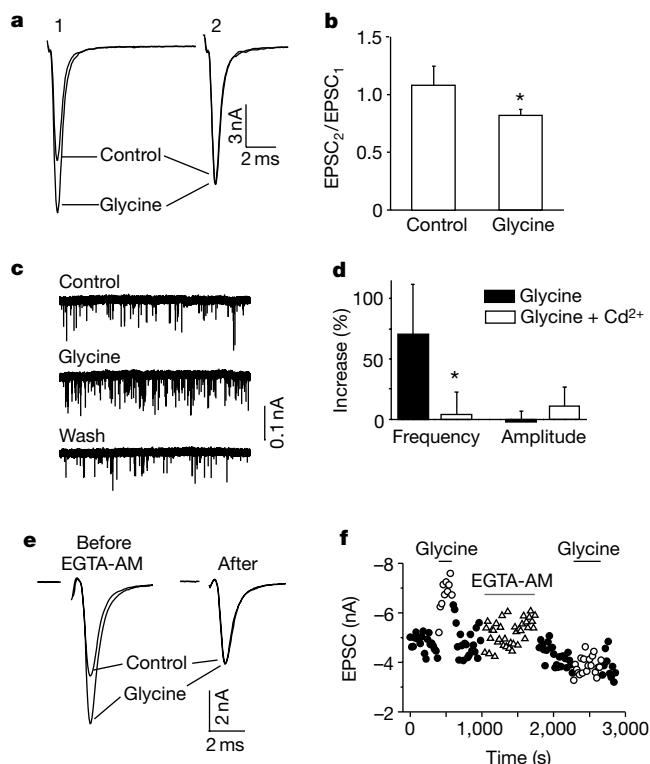


Figure 2 Presynaptic locus of glycine action. **a**, Paired responses shift from facilitation to depression during application of 1 mM glycine. Amplitude of EPSC₂ is similar in control and glycine solutions. Mean of five traces. **b**, Effects of glycine on response to paired stimuli, measured as EPSC₂/EPSC₁; asterisk, $P < 0.01$ ($n = 6$, paired t -test). **c**, One-minute-long traces showing enhancement of mEPSC frequency by 1 mM glycine. **d**, Per cent change in mean frequency and amplitude of mEPSCs with glycine; asterisk, $P < 0.01$, $n = 10$ control cells, $n = 4$ with Cd²⁺. In the presence of 100–200 μ M Cd²⁺, glycine did not alter mEPSC frequency ($P > 0.32$). **e**, EGTA-AM eliminates the effect of glycine on evoked EPSC. Averages of ten sweeps each. After incubation in EGTA-AM, control traces and traces in glycine were indistinguishable. **f**, Time course of experiment in panel **e**.

potential shift of 52.2 mV, Nernst prediction of 53.7 mV), indicating that glycine activated a Cl[−] channel (Fig. 3a–c). Similar results were obtained in all five postsynaptic recordings (not shown); thus, the glycine receptor is seemingly expressed both pre- and postsynaptically. However, to exclude the possibility that the presynaptic current evoked by glycine was influenced by glycine-induced Cl[−] flux in the postsynaptic cell, simultaneous pre- and postsynaptic recordings were made during glycine application. In these experiments, the amplitude of the presynaptic glycine response recorded at −60 mV was independent of the postsynaptic membrane potential. In three pairs, a 60-mV shift in postsynaptic membrane potential did not significantly change the amplitude of the presynaptic glycine current (−60 mV to 0 mV, $P > 0.48$, paired t -test; see Supplementary Information). Thus, the presynaptic glycine response is due to receptors localized in presynaptic membrane.

Presynaptic GABA_A receptors coupled to Cl[−] channels cause inhibition of transmitter release by reducing the presynaptic action potential and consequently reducing the Ca²⁺ current³. It was therefore surprising that glycine should activate similar channels and yet have the opposite effects on the EPSC. To explore the relation between Cl[−] current and glutamate release, we made cell-attached presynaptic recordings using gramicidin-filled patch pipettes to establish a cation-permeable perforated patch¹⁴. In current clamp, glycine produced a depolarization of about 9 mV, indicating that the intraterminal chloride concentration is normally high

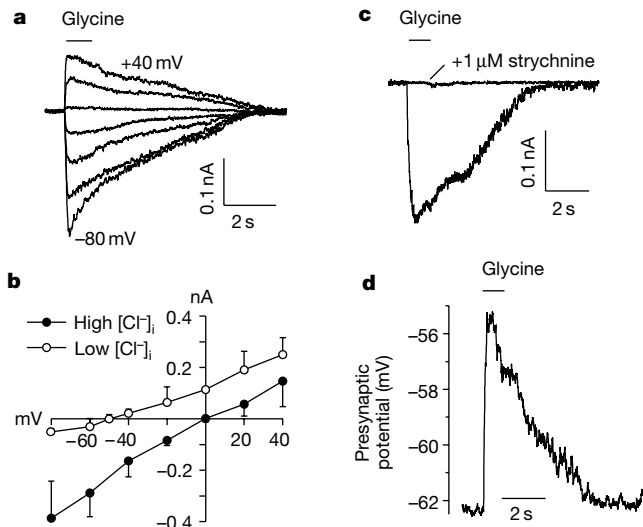


Figure 3 Glycine activates a presynaptic depolarizing Cl[−] current by glycine. **a**, Responses to 1-s application of 1 mM glycine. Cell maintained at V_{hold} from −80 mV to +40 mV in 20 mV increments. **b**, Mean current–voltage relations for glycine-evoked current for terminals filled with low (open circles, $n = 6$) or high (filled circles, $n = 6$) [Cl[−]]_i solution, as in Methods. **c**, Presynaptic response to glycine was blocked by 1 μ M strychnine. $V_{\text{hold}} = -60$ mV. **d**, Depolarizing response to 1-s pulse of 1 mM glycine during a gramicidin perforated-patch recording.

(Fig. 3d; mean 8.7 ± 3.9 mV, $n = 4$). Glycine's effects on release may therefore result from a depolarization of the terminal or, alternatively, from a metabotropic action of the receptor unrelated to the Cl[−] current.

To distinguish between these possibilities, simultaneous pre- and postsynaptic recordings were made in which a brief presynaptic current step (1 nA, 1 ms) evoked an action potential and postsynaptic EPSC. When terminals were filled with 150 mM KCl, so that glycine had a depolarizing effect (mean depolarization after 20-s application, 8.7 ± 2.9 mV), the EPSC evoked by a brief current pulse was enhanced by $60 \pm 37\%$ (Fig. 4a, $n = 4$ paired recordings, $P < 0.001$). When terminals were loaded with a low-[Cl[−]]_i solution (containing K-gluconate) so that glycine had only a minimal depolarizing effect (2.2 ± 1.4 mV), glycine did not change the EPSC amplitude (Fig. 4b, $-4 \pm 6\%$ change, $n = 4$, $P > 0.58$). However, in these same synapses, upon injection of steady, weak current through the presynaptic recording pipette to depolarize the synapses (mean 7.8 ± 3.4 mV), the amplitude of the EPSC evoked by a presynaptic action potential was significantly enhanced (Fig. 4c, four pairs with low chloride-fill, $40 \pm 11\%$; three pairs with high chloride-fill, $33 \pm 29\%$, both conditions significantly different from control, $P < 0.01$). No change was observed in the half width of the presynaptic spike upon glycine application (Fig. 4e; control, 0.54 ± 0.15 ms; glycine, 0.56 ± 0.16 ms, $n = 8$ terminals, $P < 0.8$), indicating that glycine probably does not inhibit repolarizing K⁺ currents. Given the dependence of glycine's effects on presynaptic Ca²⁺ and its sensitivity to Cd²⁺, it may be that small depolarizations activate Ca²⁺ channels, raising intracellular Ca²⁺ (ref. 15), thus resulting in potentiation of evoked release and increase in the frequency of spontaneous mEPSCs¹⁰. Consistent with this idea, six terminals loaded with the Ca²⁺ indicator Oregon Green 488 BAPTA-1 showed a gradual but significant increase in fluorescence (F) upon depolarization by 10–12 mV (Fig. 4d; mean $\Delta F/F = 0.21 \pm 0.09$, $P < 0.02$). This enhancement was prevented in three out of three terminals with 100 μ M bath Cd²⁺. These data, and the results from Cd²⁺ on mEPSCs, indicate that weak depolarization by glycine enhanced presynaptic Ca²⁺ concentrations, and thus

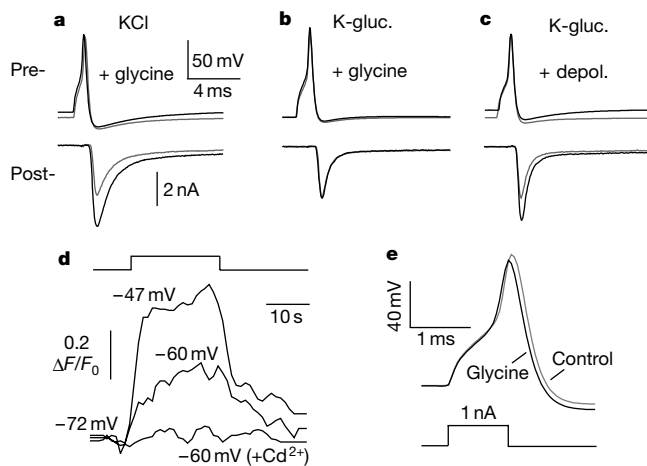


Figure 4 Glycine enhances release by depolarizing the nerve terminal. **a–c**, Simultaneous pre- and postsynaptic recordings. Upper traces are presynaptic membrane potential (current clamp) and lower traces are postsynaptic current (voltage clamp at -60 mV). Presynaptic stimulus is a 1 nA, 1 ms pulse delivered through the recording pipette. All traces are averages of three trials. **a**, KCl-filled terminal. During steady application of 1 mM glycine the terminal was depolarized and the EPSC was enhanced (control traces in grey, glycine in black). Control membrane potential -73.5 mV. **b**, K-gluconate-filled terminal. During steady application of 1 mM glycine, no depolarization or change in EPSC was evident (control traces in grey, glycine traces in black). Control membrane potential -80.7 mV. **c**, K-gluconate-filled terminal. Depolarization by weak current injection through the presynaptic patch pipette (black traces) enhanced the EPSC. Control membrane potential -78.8 mV. **d**, Change in fluorescence ($\Delta F/F_0$) in terminal loaded with Oregon Green 488 BAPTA-1 during depolarizations as indicated. Timing of depolarization shown in top trace. **e**, Presynaptic action potentials with and without glycine application in KCl-loaded terminals. Traces are baseline adjusted.

increased the likelihood of transmitter release. Such an effect might be mediated by an action of Ca^{2+} either on vesicle-release proteins or on voltage-gated Ca^{2+} channels^{16,17}.

Synaptically released glutamate and GABA are known to 'spill over' to distant presynaptic sites in cerebellum and hippocampus^{18–21}. To determine whether presynaptic glycine receptors are activated by release from glycinergic synapses, MNTB neurons of P13–15 animals were recorded at 36°C , while glutamatergic and glycinergic fibres were independently stimulated^{5,22}. A higher temperature was used to enhance glycine uptake mechanisms and thus assess the effects of endogenous glycine under more physiological conditions. Strychnine-sensitive IPSCs (inhibitory postsynaptic currents) could be generated in CsF-filled MNTB neurons held at 0 mV by using a second stimulation pipette placed near the cell body (Fig. 5a); repetitive activation of glycinergic fibres at 100 Hz resulted in a train of IPSCs (Fig. 5b). When the neurons were held at -60 mV, the IPSCs were diminished, and it was then possible to record EPSCs uncontaminated by the inhibitory current. To examine the possibility that glycine released during the IPSCs might activate receptors on the calyx, EPSCs were first evoked at -60 mV at 20 -s intervals to establish a control response. Then, each EPSC was preceded by a train of ten IPSCs at 10 -ms intervals (Fig. 5c). The EPSC obtained after a train of IPSCs was enhanced, with an average increase of $20 \pm 5\%$ (Fig. 5d, $n = 9$ cells, $P < 0.001$). Confirming that this effect was due to endogenous glycine, application of $0.3 \mu\text{M}$ strychnine blocked in the same cells both the IPSC and the enhancement of the EPSC (Fig. 5e, $n = 4$, percentage of enhancement by IPSCs before strychnine, $19 \pm 5\%$, $P < 0.03$). At this temperature, exogenous glycine (1 mM) potentiated EPSCs by $39 \pm 19\%$ ($n = 4$ cells, $P < 0.001$), roughly twice the level attained by endogenous glycine. As there are no reported axo-axonic synapses on calyx terminals, it is most likely that potentiation of

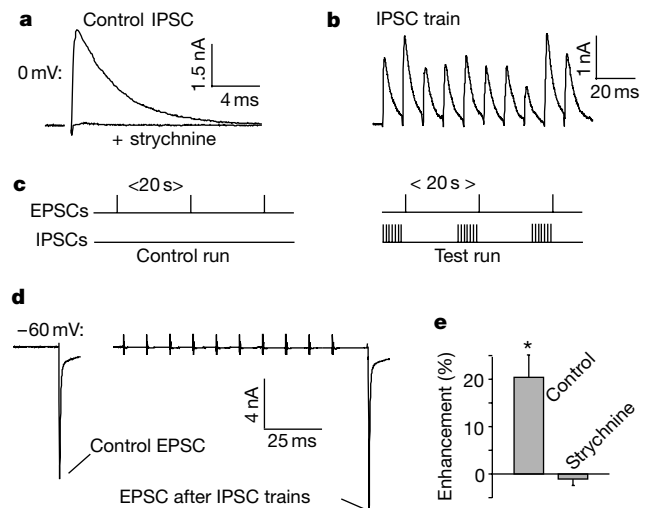


Figure 5 Spillover to presynaptic glycine receptors. **a**, A single IPSC and its block by $0.3 \mu\text{M}$ strychnine. **b**, Train of ten IPSCs used for conditioning protocol. V_{hold} for **a** and **b** was at 0 mV. **c**, Paradigm for revealing glycine spillover. EPSCs were first evoked at least ten times at 20 -s intervals to establish the control amplitude. Then, each EPSC was evoked 15 ms after the end of an IPSC train (ten shocks at 100 Hz). This test protocol was also repeated at least ten times and the amplitude of the EPSC was averaged. **d**, Example of control and test EPSCs showing enhancement by IPSC trains. **e**, Mean enhancement of EPSC by IPSCs in control and strychnine solutions; asterisk, $P < 0.001$.

EPSCs by IPSC trains results from glycine spillover from somatic boutons.

Glycine is the major inhibitory transmitter in the superior olivary complex, and is involved in the encoding of interaural intensity differences. Glycinergic neurons in MNTB mediate inhibition of the lateral superior olive by signals from the contralateral ear^{23,24}. Postsynaptic glycinergic IPSCs may shorten the membrane time constant and prevent overlap of sequential synaptic potentials evoked at high frequencies²⁵. We report several lines of evidence indicating that glycine also acts presynaptically. Glycine increased EPSC amplitude, synaptic depression and the frequency of spontaneous transmitter release, and directly depolarized nerve terminals by an amount that increases intraterminal Ca^{2+} and synaptic efficacy. During intense activation of glycinergic fibres, glycine reached presynaptic glycine receptors and increased the strength of transmission, perhaps to counter a reduction in peak EPSPs by the shunting action of postsynaptic glycine receptors. Together, the pre- and postsynaptic effects of glycine are consistent with the overall function of the MNTB, to deliver rapid, well-timed inhibition. The enhancement of release by depolarizing Cl^- channels in nerve terminals has not been previously described in the brain, despite repeated observations of presynaptic GABA_A receptors^{1–4}. However, it is possible that enhancement of transmission could occur in these other cases as well, and indeed may be a significant physiological effect of the receptor. Our results indicate that release may be enhanced by low-level activation of Cl^- channels just sufficient to depolarize, but not inactivate, large currents associated with the action potential³. A large enhancement of release as a result of a small depolarization by glycinergic receptors is consistent with recent studies showing a high sensitivity of the transmitter release machinery in MNTB to small increments in intracellular Ca^{2+} (refs 26, 27). □

Methods

Brainstem slices were prepared from P10–16 rats^{5,13} and stored in a 37°C solution containing (in mM) 125 NaCl, 25 glucose, 2.5 KCl, 1 MgCl_2 , 2 CaCl_2 , 1.25 NaH_2PO_4 , 25 NaHCO_3 , 0.4 ascorbic acid, 3 *myo*-inositol and 2 sodium pyruvate, bubbled with 5% $\text{CO}_2/95\%$ O_2 . During recordings (21 – 23°C , unless indicated otherwise), slices were perfused

with either the solution listed above (without ascorbic acid, *myo*-inositol, and Na-pyruvate) or a HEPES-based solution containing (in mM) 145 NaCl, 25 glucose, 2.5 KCl, 1 MgCl₂, 2 CaCl₂, 10 HEPES, bubbled with O₂ (used for Fig. 3a–c). During spillover experiments, the extracellular solution also contained 60 nM CGP54626 to prevent activation of GABA_A receptors. Presynaptic terminals were labelled with lucifer yellow and visually identified as a calyx.

Electrodes for postsynaptic recording had resistances of 2–3 MΩ; series resistances during recordings were <5 MΩ, compensated electronically by 90%. Pre- and postsynaptic cells were voltage clamped to –60 mV, unless otherwise indicated. For presynaptic recordings, electrodes were 5–6 MΩ, with series resistances of 10–25 MΩ, compensated by 80–90%. Pipettes for recording EPSCs contained (in mM) 135 CsF, 5 CsCl, 5 EGTA, 10 HEPES and 2 QX314 (*N*-(2,6-dimethylphenylcarbamoylmethyl) triethylammonium chloride) (286 mOsm) at pH 7.25 with CsOH. Pipettes for pre- and postsynaptic recording of glycine responses contained either low [Cl[–]]_i solution (in mM): 125 Cs-methane sulphonate, 15 CsCl, 5 EGTA, 1 MgCl₂, 10 HEPES, 0.2 lucifer yellow (290 mOsm) at pH 7.2 with CsOH, or high-[Cl[–]]_i solution in which Cs-methane sulphonate was replaced with CsCl. Current-clamp recording from gramicidin-perforated patches was made with pipettes filled with high-[Cl[–]]_i solution where KCl replaced Cs-methane sulphonate and CsCl. For paired pre- and postsynaptic recordings, presynaptic pipettes contained (in mM) 150 KCl (or 145 K-gluconate and 5 KCl for low-[Cl[–]]_i solution), 0.2 EGTA, 1 MgCl₂ and 10 HEPES¹³. During simultaneous pre- and postsynaptic recordings, no rundown of EPSCs was observed over the 10–15 min necessary for the experiment. Voltages were corrected for junction potentials.

EPSCs and IPSCs were elicited by voltage pulses (100 μs, 10–20 V stimuli) delivered through a glass pipette. During paired pre- and postsynaptic recording, EPSCs were evoked by action potentials elicited by a depolarizing current pulse delivered through the presynaptic recording electrode. mEPSCs were recorded at –60 mV in the presence of 0.5 μM TTX. Drugs were applied by pressure ejection or by bath perfusion. Gramicidin D was prepared in methanol at 5 mg ml^{–1} and then dissolved in intracellular solution to a final concentration of 5 μg ml^{–1}. EGTA-AM prepared as a stock solution in dimethyl sulphoxide (DMSO) just before the experiment was diluted in bath solution to 0.2 mM (final concentration of DMSO, 0.1%). For Ca²⁺ measurements, voltage-clamped terminals were loaded with the presynaptic K-gluconate pipette fill containing 20 μM Oregon Green 488 BAPTA-1, and without Mg²⁺. Dye was excited (Hg lamp with 1% neutral density filter, 470-nm bandpass excitation filter) at 1 Hz for 100 ms to minimize bleaching. Fluorescence emission (535 nm bandpass) was detected with a photomultiplier tube and was restricted to a region just larger than the cell body diameter. Background values, measured in a region of identical size adjacent to cell body, were subtracted and fluorescence changes normalized for the average pre-stimulus intensity ($\Delta F/F_0$). Statistical significance was established using paired and unpaired *t*-tests, as indicated, and errors reported as ± 1 s.d.

Received 24 November 2000; accepted 27 March 2001.

- Eccles, J. C., Schmidt, R. F. & Willis, W. D. Pharmacological studies of presynaptic inhibition. *J. Physiol.* **168**, 500–530 (1963).
- Tachibana, M. & Kaneko, A. γ -Aminobutyric acid exerts a local inhibitory action on the axon terminal of bipolar cells: evidence for negative feedback from amacrine cells. *Proc. Natl Acad. Sci. USA* **84**, 3501–3505 (1987).
- Zhang, S. J. & Jackson, M. B. GABA_A receptor activation and the excitability of nerve terminals in the rat posterior pituitary. *J. Physiol. (Lond.)* **483**, 583–595 (1995).
- Pouzat, C. & Marty, A. Somatic recording of GABAergic autoreceptor current in cerebellar stellate and basket cells. *J. Neurosci.* **19**, 1675–1690 (1999).
- Forsythe, I. D. & Barnes-Davies, M. The binocular auditory pathway: excitatory amino acid receptors mediate dual timecourse excitatory postsynaptic currents in the rat medial nucleus of the trapezoid body. *Proc. R. Soc. Lond. B.* **25**, 151–157 (1993).
- Taschenberger, H. & von Gersdorff, H. Fine tuning an auditory synapses for speed and fidelity during a critical developmental period. *J. Neurosci.* **24**, 9162–9173 (2000).
- Bormann, J., Hamill, O. P. & Sakmann, B. Mechanism of anion permeation through channels gated by glycine and gamma-aminobutyric acid in mouse cultured spinal neurones. *J. Physiol. (Lond.)* **385**, 243–286 (1987).
- Isaacson, J. S. GABA_B receptor-mediated modulation of presynaptic currents and excitatory transmission at a fast central synapse. *J. Neurophysiol.* **80**, 1571–1576 (1998).
- Takahashi, T., Kajikawa, Y. & Tsujimoto, T. G-Protein-coupled modulation of presynaptic calcium currents and transmitter release by a GABA_A receptor. *J. Neurosci.* **18**, 3138–3146 (1998).
- Martin, A. R. In *Handbook of Physiology, Sec. I, The Nervous System* (eds Brookhard, J. M., Mountcastle, V. B., Kandel, E. R. & Geiger, S. R.) 329–356 (Am. Physiol. Soc., Bethesda, 1977).
- Atluri, P. P. & Regehr, W. G. Delayed release of neurotransmitter from cerebellar granule cells. *J. Neurosci.* **18**, 8214–8227 (1998).
- Wang, L. Y. & Kaczmarek, L. K. High-frequency firing helps replenish the readily releasable pool of synaptic vesicles. *Nature* **394**, 384–388 (1998).
- Borst, J. G. & Sakmann, B. Calcium influx and transmitter release in a fast CNS synapse. *Nature* **383**, 431–434 (1996).
- Akaike, N. Gramicidin perforated patch recording and intracellular chloride activity in excitable cells. *Prog. Biophys. Mol. Biol.* **65**, 251–264 (1996).
- Tachibana, M., Okada, T., Arimura, T., Kobayashi, K. & Piccolino, M. Dihydropyridine-sensitive calcium current mediates neurotransmitter release from bipolar cells of the goldfish retina. *J. Neurosci.* **13**, 2898–2909 (1993).
- Cuttle, M. F., Tsujimoto, T., Forsythe, I. D. & Takahashi, T. Facilitation of the presynaptic calcium current at an auditory synapse in rat brainstem. *J. Physiol. (Lond.)* **512**, 723–729 (1998).
- Borst, J. G. & Sakmann, B. Facilitation of presynaptic calcium currents in the rat brainstem. *J. Physiol. (Lond.)* **513**, 149–155 (1998).
- Isaacson, J. S., Solis, J. M. & Nicoll, R. A. Local and diffuse synaptic actions of GABA in the hippocampus. *Neuron* **10**, 165–175 (1993).
- Dittman, J. S. & Regehr, W. G. Mechanism and kinetics of heterosynaptic depression at a cerebellar synapse. *J. Neurosci.* **17**, 9048–9059 (1997).

- Min, M. Y., Rusakov, D. A. & Kullmann, D. M. Activation of AMPA, kainate, and metabotropic receptors at hippocampal mossy fiber synapses: role of glutamate diffusion. *Neuron* **21**, 561–570 (1998).
- Mitchell, S. J. & Silver, R. A. Glutamate spillover suppresses inhibition by activating presynaptic mGluRs. *Nature* **404**, 498–502 (2000).
- Banks, M. I. & Smith, P. H. Intracellular recordings from neurobiotin-labeled cells in brain slices of the rat medial nucleus of the trapezoid body. *J. Neurosci.* **12**, 2819–2837 (1992).
- Wu, S. H. & Kelly, J. B. Synaptic pharmacology of the superior olivary complex studied in mouse brain slice. *J. Neurosci.* **12**, 3084–3097 (1992).
- Grothe, B. & Sanes, D. H. Synaptic inhibition influences the temporal coding properties of medial superior olivary neurons: an in vitro study. *J. Neurosci.* **14**, 1701–1709 (1994).
- Funabiki, K., Koyano, K. & Ohmori, H. The role of GABAergic inputs for coincidence detection in the neurones of nucleus laminaris of the chick. *J. Physiol. (Lond.)* **508**, 851–869 (1998).
- Schneggenburger, R. & Neher, E. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature* **406**, 889–893 (2000).
- Bollmann, J. H., Sakmann, B. & Borst, J. G. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science* **289**, 953–957 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Brenowitz, H. von Gersdorff and C. Jahr for comments on the manuscript. Supported by NIH grants and a Fogarty International Center Fellowship.

Correspondence and requests for materials should be addressed to L.T. (e-mail: trussell@ohsu.edu).

LTRPC7 is a Mg·ATP-regulated divalent cation channel required for cell viability

Monica J. S. Nadler[†], Meredith C. Hermosura[‡], Kazunori Inabe[§], Anne-Laure Perraud[†], Qiqin Zhu[†], Alexander J. Stokes[†], Tomohiro Kurosaki[§], Jean-Pierre Kinet[†], Reinhold Penner[‡], Andrew M. Scharenberg^{†||} & Andrea Fleig[‡]

[†] Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA

[‡] Laboratory of Cell and Molecular Signaling, Center for Biomedical Research at The Queen's Medical Center and John A. Burns School of Medicine at the University of Hawaii, Honolulu, Hawaii 96813, USA

[§] Department of Molecular Genetics, Institute for Liver Research, Kansai Medical University, Osaka 570-8506, Japan

* These authors contributed equally to this work

The molecular mechanisms that regulate basal or background entry of divalent cations into mammalian cells are poorly understood. Here we describe the cloning and functional characterization of a Ca²⁺- and Mg²⁺-permeable divalent cation channel, LTRPC7 (nomenclature compatible with that proposed in ref. 1), a new member of the LTRPC family of putative ion channels. Targeted deletion of LTRPC7 in DT-40 B cells was lethal, indicating that LTRPC7 has a fundamental and nonredundant role in cellular physiology. Electrophysiological analysis of HEK-293 cells overexpressing recombinant LTRPC7 showed large currents regulated by millimolar levels of intracellular Mg·ATP and Mg·GTP with the permeation properties of a voltage-independent divalent cation influx pathway. Analysis of several cultured cell types demonstrated small magnesium-nucleotide-regulated metal ion currents (MagNuM) with regulation and permeation properties essentially identical to the large currents observed in cells expressing recombinant LTRPC7. Our data indicate that LTRPC7, by

|| Present address: Department of Pediatrics and Immunology, University of Washington and Children's Hospital and Medical Center, Seattle, Washington 98195-6320, USA.

virtue of its sensitivity to physiological Mg-ATP levels, may be involved in a fundamental process that adjusts plasma membrane divalent cation fluxes according to the metabolic state of the cell.

As part of a systematic effort to identify novel Ca^{2+} /cation channels expressed in haematopoietic cells, we cloned a putative cation channel transcript, which we later designated LTRPC7. The predicted protein encoded by the LTRPC7 transcript is homologous to other LTRPC family members throughout its first 1,200 amino acids (Fig. 1a), but is notable for its ubiquitous distribution (see Supplementary Information) and for a long and unique carboxy-terminal tail that contains a domain of significant homology to the MHCK/EEF2 α family of protein kinases² (Fig. 1a).

To investigate LTRPC7's role in cell physiology, we used the DT-40 B-cell system to produce a cell line with targeted deletion of its LTRPC7 genes. Although targeted deletion of one LTRPC7 allele was easily obtained, no clones were obtained with targeted integrations into both alleles (data not shown), indicating that LTRPC7 is required for cellular viability. We therefore took an inducible deletion approach (Fig. 1b) by producing a viable DT-40 cell line that expressed a tamoxifen-controlled Cre recombinase, and which also had a stable targeted deletion of one LTRPC7 allele and LoxP sites bracketing the LTRPC7 exons encoding transmembrane span 5, the putative channel pore region and transmembrane span 6 in the other LTRPC7 allele. Activation of the Cre recombinase by treatment of this cell line with tamoxifen induced the deletion of these three exons in most cells (Fig. 1c). As shown in Fig. 1d, this

induced the growth arrest and subsequent death of the cells over a 48–72-h time frame, providing strong evidence that LTRPC7 is fundamental in cellular function.

For electrophysiological analysis of LTRPC7, we produced HEK-293 cell lines with tetracycline-controlled expression of a Flag-tagged LTRPC7 construct. Treatment of these cells with tetracycline for 24 h induced a Flag-reactive band with a relative molecular mass of 220,000 (M_r 220K; predicted size of Flag-LTRPC7, Fig. 2a). However, by 72–96 h, LTRPC7-expressing cells began to swell, detach and die (Fig. 2b, similar results for all clones). Because of the potential for cell toxicity at times greater than 24 h after induction, all subsequent analyses were performed 18–24 h after induction. In this time frame, induced cells developed large whole-cell currents over a 100–300-s time period (Fig. 2c) with a characteristic outwardly rectifying current–voltage relationship (Fig. 2d) when perfused with Cs^+ -based internal solutions (identical results were obtained with K^+ , data not shown). With choline-based intracellular solutions, the large outward conductances were significantly suppressed, indicating that outward currents through LTRPC7 are carried by cations. Stationary noise analysis of outward currents was performed at +60 mV in the frequency range 0.001–1 kHz and revealed characteristic increases in current variance during activation of the whole-cell current that levelled off and finally decreased as the current was fully activated. Through linear regression of variance versus mean current during the initial activation phase, we determined an average slope of 2.5 ± 0.4 pA ($n = 4$), corresponding to a single-channel conductance of ~ 40 pS, based on a reversal potential of 0 mV. Because mono- and divalent ions do not permeate LTRPC7 independently of each other (see below), this value applies in a strict manner only to the test potential of +60 mV and the ionic conditions imposed by our external and internal solutions.

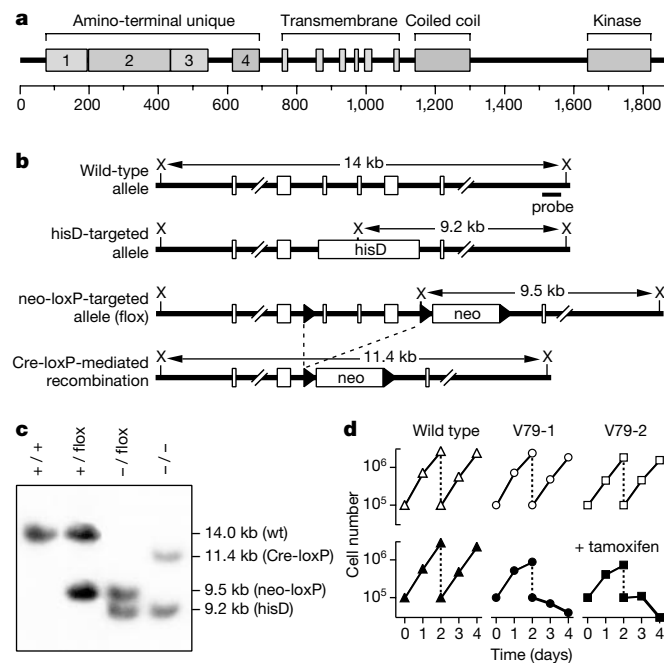


Figure 1 LTRPC7 structure and inducible knockout of LTRPC7 in the chicken B-cell line DT-40. **a**, Schematic of LTRPC7 with amino-terminal unique regions 1–4, transmembrane domain regions, coiled-coil region and the MHCK/EEF2 α kinase homology domain (kinase). **b**, Schematics of wild-type and mutated DT-40 LTRPC7 alleles. Restriction enzyme sites (X, *Xba*I), probe for Southern blot analysis (solid bar), exons (open rectangle) and loxP sites (solid triangle) are indicated. Three exons, including part of the putative transmembrane region (corresponding to mouse LTRPC7 amino-acid residues 997–1,158), were replaced with *hisD* cassettes in the *hisD*-targeted allele and flanked by two loxP sequences in the *neo*-loxP-targeted allele. *Xba*I fragments detected by the probe are shown for wild-type and mutated alleles. **c**, Southern blot analysis of *Xba*I-digested DNA from wild-type and mutant DT-40 cells. **d**, Effect of LTRPC7 inactivation on cell proliferation. DT-40 wild type and mutant clones (V79-1 and V79-2) harbouring *hisD*- and *neo*-loxP-targeted alleles were cultured either without (open symbols) or with (filled symbols) 200 nM tamoxifen. Cell numbers were adjusted to 1×10^5 cells per ml, 2 days after cultivation. Viable cells were monitored daily by trypan blue exclusion.

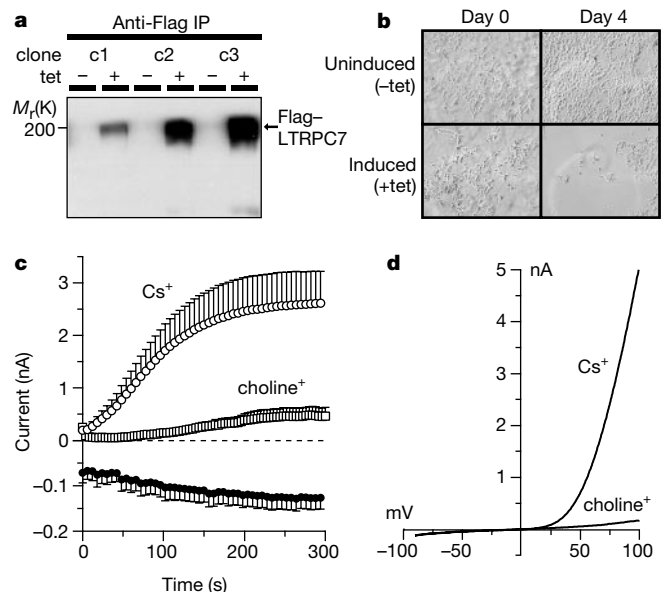


Figure 2 Functional expression of LTRPC7 as a cation channel. **a**, SDS–polyacrylamide gel electrophoresis (PAGE) analysis of three clonal HEK-293 cell lines (c1–c3) expressing a tetracycline-controlled Flag–LTRPC7 construct. Anti-Flag immunoreactive proteins before (left lanes) or after 24 h of tetracycline (right lanes). **b**, Phase contrast images of representative patches of LTRPC7-overexpressing HEK-293 cells 0 and 4 days after treatment with or without tetracycline. **c**, Average inward and outward currents carried by LTRPC7 at –80 and +80 mV, respectively. Cells perfused with ATP-free Cs^+ -glutamate internal solution ($n = 7$, $\pm$ s.e.m.). Outward current was markedly suppressed when choline was used as the major internal cation, with only 10 mM Cs^+ present ($n = 3$, $\pm$ s.e.m.). **d**, Current–voltage relationships under experimental conditions as in **c**, obtained from representative cells 200 s after whole-cell establishment.

We next analysed the permeation characteristics of LTRPC7-mediated currents (Fig. 3). Inward currents were not affected by substitution of choline for Na^+ and K^+ in the extracellular solution (Fig. 3a), indicating that the inward current is carried exclusively by divalent ions. Furthermore, inward currents are not affected by removal of either Ca^{2+} or Mg^{2+} alone (Fig. 3b, c). Only the complete removal of external divalents amplifies both inward and outward monovalent current flow (Fig. 3d). The current–voltage relationship under divalent-free conditions (Fig. 3e) demonstrates linearization of the outward currents and significant enhancement of inward currents as compared with the divalent-containing situation. These effects are probably due to relief from permeation block by the divalent ions, which is expected to be strongest at negative potentials and gradually decreases as the reversal potential for divalent ions is approached.

The above data are most simply explained by a model in which the LTRPC7 pore has a high affinity for and is permeant to Ca^{2+} and Mg^{2+} , which accordingly obstruct inward fluxes of monovalent cations. Consistent with this model, LTRPC7-dependent inward currents are enhanced and outward currents suppressed when cells are exposed to isotonic (120 mM) CaCl_2 and MgCl_2 solutions, conditions in which Ca^{2+} or Mg^{2+} are the only available cationic charge carriers (Fig. 3f and g, respectively). In light of the significant Mg^{2+} permeation of LTRPC7, a feature to our knowledge unprecedented among known ion channels, we assessed Mg^{2+} permeation at different extracellular Mg^{2+} concentrations in reference to inward currents carried by 2 mM Mg^{2+} (Fig. 3h; note the absence of external Ca^{2+}). Higher concentrations of Mg^{2+} increased inward currents carried by Mg^{2+} and submillimolar Mg^{2+} concentrations progressively increased monovalent influx until divalent-free medium gave rise to large monovalent inward currents. Thus, LTRPC7 exhibits clear Mg^{2+} -dependent anomalous mole fraction behaviour. At the same time, outward currents are inhibited with increasing Mg^{2+}

concentrations (Fig. 3i), probably in part owing to the shift in reversal potential occurring with the increase in extracellular Mg^{2+} concentrations, but also indicating that Mg^{2+} upon permeation might suppress LTRPC7 intracellularly (see below).

The time course of passive activation of currents carried by LTRPC7 is most consistent with activation due to diffusional depletion of a small intracellular constituent that suppresses channel opening. On the basis of the results above and the presence of the α -kinase domain in LTRPC7, we suspected that $[\text{Mg}^{2+}]_i$ and $[\text{ATP}]_i$ might be involved in LTRPC7 regulation. We therefore systematically changed the $[\text{Mg}^{2+}]_i$ and $[\text{ATP}]_i$ of our patch pipette solutions. LTRPC7 channel activity is strongly suppressed by $\text{Mg}\cdot\text{ATP}$ concentrations in the millimolar range (Fig. 4a, d), with nearly complete suppression of currents at 6 mM $\text{Mg}\cdot\text{ATP}$. Under these conditions, free $[\text{Mg}^{2+}]_i$ varies from 670 to 800 μM , and the available ATP is essentially all in the physiological $\text{Mg}\cdot\text{ATP}$ form. However, no suppressive effect is observed when 2 mM $\text{Na}\cdot\text{ATP}$ is added to a $[\text{Mg}^{2+}]_i$ -free internal solution, weighing against a role for a standard high-energy phosphate group transfer in the suppression phenomenon as well as indicating that $[\text{Mg}^{2+}]_i$ is a required factor or cofactor for inhibition. Figure 4b summarizes the average currents observed during intracellular perfusion of cells with solutions devoid of ATP but containing different free Mg^{2+} concentrations. From this, it is evident that high $[\text{Mg}^{2+}]_i$ strongly suppresses LTRPC7-dependent currents even in the absence of ATP, with 3 mM causing a complete block (Fig. 4e).

To evaluate further the potential role of high-energy phosphate transfer in suppression, we perfused cells with four different Mg^{2+} -nucleotide triphosphates (NTP) to define the nucleotide specificity (Fig. 4c). At 2 mM (free $[\text{Mg}^{2+}]_i \approx 700\text{--}850 \mu\text{M}$), both ATP and GTP largely suppressed activation of LTRPC7 currents (with essentially identical dose–response curves), whereas CTP and ITP were less effective compared with NTP-free controls. If the dose–

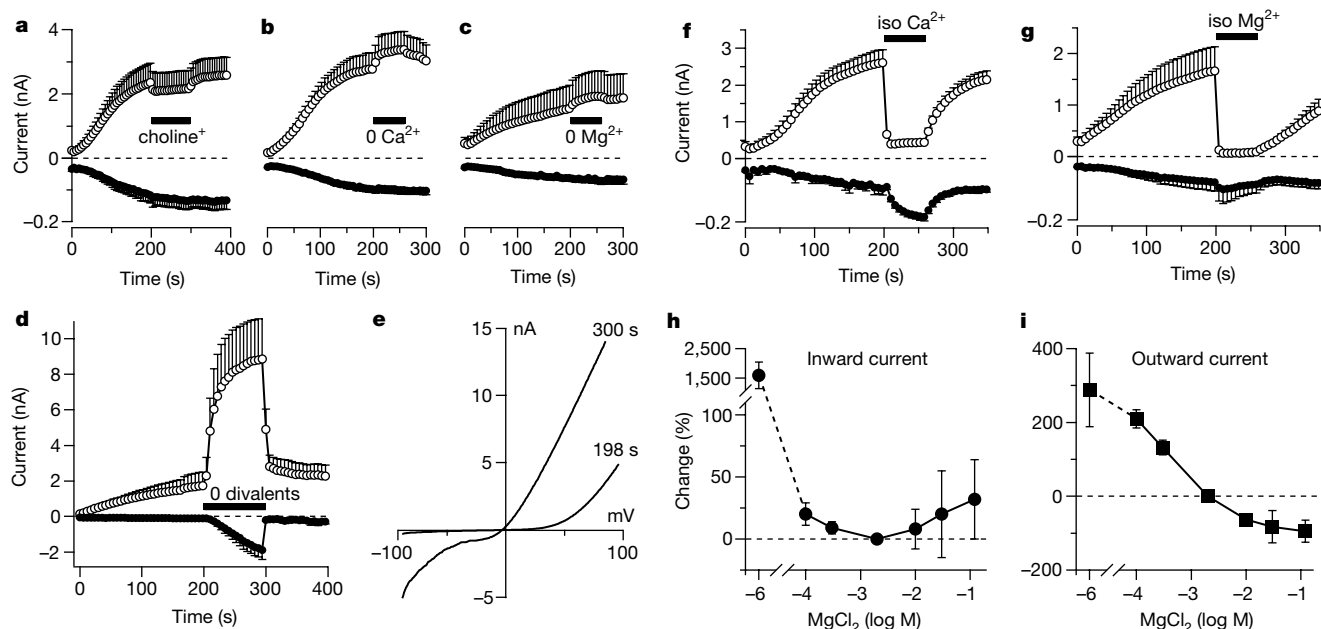


Figure 3 Permeation and block of LTRPC7 by divalent ions. Average inward and outward currents carried by recombinant LTRPC7 at -80 and $+80$ mV, respectively. Black bars indicate solution changes. **a**, Extracellular solutions contained 10 mM Ca^{2+} and 2 mM Mg^{2+} . Inward currents remained unaffected by application of extracellular solution in which NaCl was replaced by choline-Cl ($n = 5$). **b**, Application of Ca^{2+} -free extracellular solution ($n = 5$). **c**, Application of Mg^{2+} -free extracellular solution ($n = 5$). **d**, Removal of both Ca^{2+} and Mg^{2+} ($n = 5$). **e**, I – V relationship of currents under conditions as in **d** elicited before divalent-free application (198 s) and just before re-admission of divalents

(300 s). **f**, **g**, Application of isotonic CaCl_2 or MgCl_2 (120 mM) reversibly enhanced inward currents and strongly inhibited outward currents ($n = 5$ each). **h**, Anomalous mole fraction behaviour of Mg^{2+} permeation. Changes in inward current measured at -80 mV as a function of extracellular Mg^{2+} concentration ($n = 3\text{--}5$; Mg^{2+} -free solution plotted at $1 \mu\text{M}$) and normalized in reference to 2 mM extracellular Mg^{2+} (extracellular solutions were nominally Ca^{2+} -free). **i**, Analysis of the corresponding changes in outward currents (measured at $+80$ mV) in the same cells as in **g** (also in reference to 2 mM Mg^{2+}).

response curve for ATP (Fig. 4d) is instead plotted as a function of calculated free $[Mg^{2+}]_i$ (Fig. 4e, open boxes), it can be seen that suppression of LTRPC7 currents by Mg-ATP occurs within a range of free $[Mg^{2+}]_i$ that produces only moderate suppressive effects on LTRPC7 in the absence of ATP or GTP (Fig. 4e, filled circles). Together, these results would be most consistent with Mg-ATP acting as a physiological regulator of LTRPC7, as cellular levels of free $[Mg^{2+}]_i$ are around 0.5 mM (ref. 3) and cytosolic GTP levels are at least five times lower than those of ATP, which are typically several millimolar⁴.

Finally, as a further test of a role for a high-energy phosphate transfer reaction in LTRPC7 gating, we performed experiments with Na-ATP- γ -S and Mg-ATP- γ -S, non-hydrolysable analogues of ATP. Because ATP- γ -S can be used by protein kinases in thio-phosphorylation reactions, and thio-phosphate linkages are not susceptible to protein phosphatases, the result is irreversible protein phosphorylation. As shown in Fig. 4f (open diamonds), 2 mM Mg-ATP- γ -S (free $[Mg^{2+}]_i \approx 700 \mu M$) effectively suppresses LTRPC7 currents. To address the question of whether this suppression is due to irreversible thio-phosphorylation of LTRPC7, we performed experiments in which cells were initially patched with pipette solutions containing Mg-ATP- γ -S (filled circles in Fig. 4f) and subsequently re-patched around 180 s later, this time with an ATP-free pipette solution to wash out Mg-ATP- γ -S. This resulted in essentially full re-development of LTRPC7 currents, demonstrating the reversibility of the suppression phenomenon under conditions in which the participation of a phosphorylation event should have resulted in irreversible suppression. Therefore, the suppression of LTRPC7 activation by Mg-ATP- γ -S is analogous to that of Mg-ATP or Mg-GTP and is likely to be the result of direct regulation of channel

activity. By the same logic, Na-ATP- γ -S should behave similarly to Na-ATP and this is confirmed in Fig. 4f (open circles), where 2 mM Na-ATP- γ -S failed to suppress LTRPC7 activation. We did observe a Na-ATP- γ -S-mediated delayed and slow inactivation, which may involve phosphorylation events that downregulate LTRPC7 activity.

We also analysed endogenous currents activated in several cultured cell lines from tissues positive for LTRPC7 transcripts, using the same experimental conditions that produced maximal LTRPC7 activation in our heterologous HEK-293 system. Each cell type analysed (Fig. 5) possessed currents with Mg-ATP sensitivity and current-voltage relationship similar to those observed in HEK-293 cells overexpressing LTRPC7. In particular, the native currents exhibited strong outward rectification at potentials above +50 mV, which may be considered the signature of this conductance. We have adopted the convention that native currents that exhibit sensitivity to Mg-nucleotide complexes and permeation properties like those of LTRPC7 are designated as magnesium-nucleotide-regulated metal ion currents or MagNum. Our data indicate that LTRPC7 mediates most or all of MagNum in all cell types so far analysed.

An analysis of LTRPC7 function has also been performed by Clapham and colleagues⁵. They concluded that the EEF2 α domain kinase activity was required for channel gating. However, this conclusion was based in part on increased currents observed upon addition of 5 mM Na-ATP to a standard intracellular solution containing only 1 mM $MgCl_2$, reducing free $[Mg^{2+}]_i$ to 6 μM . As our data show Mg^{2+} to be an essential (co)factor in regulating LTRPC7 activity, such a drop in free $[Mg^{2+}]_i$ would be expected to produce a large relative increase in current amplitude independent of ATP, suggesting that conclusions regarding the role of the kinase domain in channel activation should be taken cautiously. Clapham and

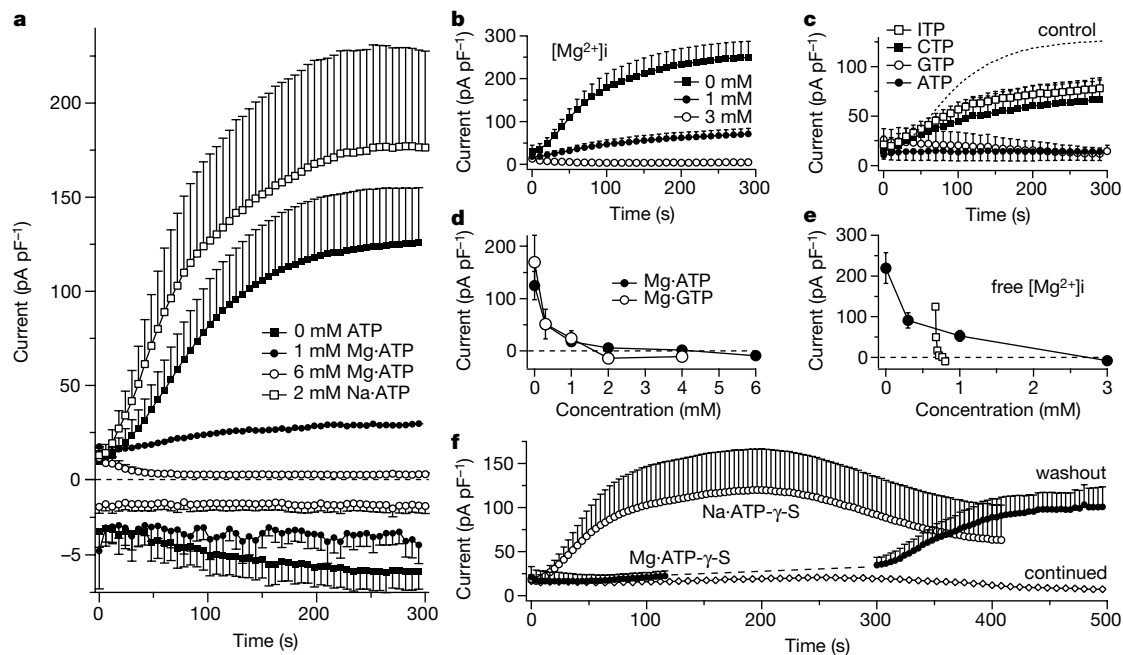


Figure 4 LTRPC7 gating is modulated by $[Mg^{2+}]_i$ and Mg-nucleotides. **a**, Average inward and/or outward currents carried by recombinant LTRPC7 at -80 and $+80$ mV, respectively. Cells were perfused with internal solutions containing various ATP concentrations (0 mM Mg-ATP, $n = 7$; 1 mM Mg-ATP, $n = 5$; 6 mM Mg-ATP, $n = 5$; 2 mM Na-ATP, $n = 5$). **b**, Cells were perfused with internal solutions containing the indicated $MgCl_2$ concentrations in the absence of added ATP (in all cases $n = 5$). **c**, Cells were perfused with internal solutions containing 2 mM of various Mg-nucleotides (in all cases $n = 5$). Dotted control trace represents 0 ATP taken from **a**. **d**, Average changes of maximum outward current measured at $+80$ mV as a function of intra-pipette Mg-ATP or Mg-GTP levels. The change in current size was analysed by subtracting the first data trace acquired after whole-cell establishment from the one elicited at 300 s. **e**, Average

changes of maximum outward current measured at $+80$ mV as a function of intra-pipette free Mg^{2+} levels in the absence of added ATP (filled circles). The change in current size was analysed as in **d**. Open squares, data points of the ATP dose-response curve in **d**, here plotted against the calculated free Mg^{2+} under those conditions. **f**, Cells were perfused with internal solutions containing 2 mM of various non-hydrolysable analogues of ATP (Na-ATP- γ -S, open circles, $n = 6$; Mg-ATP- γ -S, open diamonds, $n = 3$). Filled circles represent three cells perfused with 2 mM Mg-ATP- γ -S for 120 s and then re-patched with ATP-free solution. Data sets from re-patched cells were aligned, averaged and plotted with a delay of 180 s, which represents the average time needed to exchange pipettes and re-patch individual cells (120 s, 198 s, 215 s, respectively).

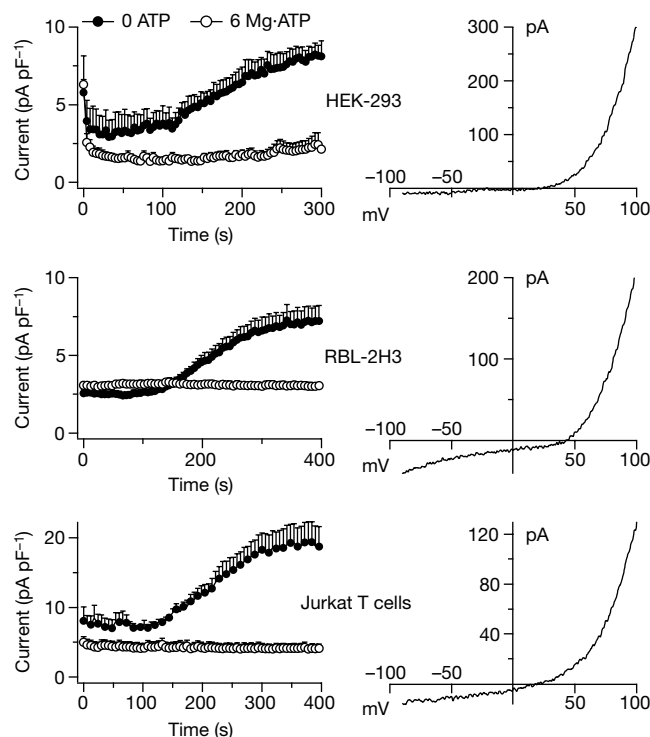


Figure 5 Mg-ATP sensitive conductances with LTRPC7 signature are ubiquitous. Wild-type HEK-293, RBL-2H3 and Jurkat T-lymphocytes were perfused with standard internal solutions supplemented with either 6 mM ATP (open circles; $n = 5$, $\pm$ s.e.m.) or no ATP

colleagues also concluded, on the basis of reversal potential measurements, that LTRPC7 is essentially non-selective for monovalent relative to divalent cations. However, as monovalent and divalent ions do not permeate LTRPC7 independently (Fig. 3), the permeation of LTRPC7 at physiological membrane potentials (-40 to -80 mV) is selective for divalent ions, whereas monovalent permeation becomes significant only at rather positive potentials.

In conclusion, we have cloned a member of the LTRPC family of ion channels, designated LTRPC7, and analysed its role in cellular function. Our data indicate that LTRPC7 is an intracellular ligand-gated ion channel whose activation may be linked to cellular energy metabolism through its sensitivity to cytosolic Mg-ATP levels, and that it effectively permeates both Ca^{2+} and Mg^{2+} . As loss of LTRPC7 function would presumably disrupt the coordination of cellular energy production with the homeostasis of these (and possibly other) divalent cations, these functional characteristics plausibly account for the cellular death that occurs in standard cell culture conditions upon deletion of the DT-40 LTRPC7 genes. Furthermore, because of its link to cellular energy metabolism, LTRPC7 may represent a direct mediator of the toxic divalent cation entry that occurs during severe metabolic stress conditions such as hypoxia or hypoglycaemia. □

Methods

Expression analysis

RT-PCR (PCR with reverse transcription) analysis was performed from the indicated human tissue complementary DNA libraries according to the manufacturer's protocols (Life Technologies). For LTRPC7, oligonucleotides used were GTCACCTGGAACTGG AACC and CGGTAGATGGCCTTCTACTG to produce a 278-base-pair (bp) band. PCR was performed using standard techniques and 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 60 s. The approximate intensity of the ethidium bromide staining of correct sized bands was estimated by eye to be 1–2+. The LTRPC7 primers used in these reactions were generated from initial expressed sequence tag sequences, and contain a single base pair mismatch at the 5' end of the primer based on the corresponding region of LTRPC7 sequence obtained from subsequent clones. Multiple tissue northern blots for human tissues and cell lines were obtained from Clontech and all hybridizations were performed

(filled circles; $n = 5$, $\pm$ s.e.m.). Under these experimental conditions the store-operated Ca^{2+} current I_{CRAC} contributes a larger proportion of inward currents than when LTRPC7 is overexpressed, accounting for the right-shift in reversal potential.

according to the manufacturer's protocols. LTRPC7 northern blots were performed using a dUTP-labelled RNA probe generated from a 500-bp fragment corresponding to approximately residues 1,695–2,195 of the human LTRPC7 transcript. The probe was generated using a T7-directed RNA probe synthesis kit from Ambion.

Electrophysiology

HEK-293 cells transfected with the Flag-murineLTRPC7/pCDNA4-TO construct were grown on glass coverslips with DMEM supplemented with 10% fetal bovine serum, blasticidin ($5 \mu\text{g ml}^{-1}$), and zeocin (0.4 mg ml^{-1}). LTRPC7 expression was induced by adding $1 \mu\text{g ml}^{-1}$ tetracycline to the culture medium. Whole-cell patch-clamp experiments were performed at 21 – 25°C 18–24 h after induction, using cells grown on glass coverslips and kept in a standard modified Ringer's solution of the following composition (in mM): NaCl 145, KCl 2.8, CaCl_2 10, CaCl_2 1, MgCl_2 2, glucose 10, HEPES-NaOH 10, pH 7.2. In some experiments, nominally Ca^{2+} and/or Mg^{2+} -free extracellular solutions or isotonic CaCl_2 and MgCl_2 solutions (120 mM) were applied by pressure ejection from wide-tipped pipettes. Intracellular pipette-filling solutions contained (in mM): Cs-glutamate 145, NaCl 8, MgCl_2 1, Cs-BAPTA 10, HEPES-CsOH 10, pH 7.2. In some experiments, Cs-glutamate was replaced equimolarly by K-glutamate or choline-chloride. Free $[\text{Mg}^{2+}]$ was calculated by Patcher's Power Tools (<http://www.wavemetrics.com/Users/ppt.html>). High-resolution current recordings were acquired by a computer-based patch-clamp amplifier system (EPC-9, HEKA). Immediately following establishment of the whole-cell configuration, voltage ramps of 50-ms duration spanning the voltage range of -100 to $+100$ mV were delivered from a holding potential of 0 mV at a rate of 0.5 Hz over a period of 200–400 s. All voltages were corrected for a liquid junction potential of 10 mV between external and internal solutions when internal solutions contained glutamate. Currents were filtered at 2.3 kHz and digitized at 100- μs intervals. Capacitive currents and series resistance were determined and corrected before each voltage ramp using the automatic capacitance compensation of the EPC-9. The low-resolution temporal development of currents at a given potential was extracted from individual ramp current records by measuring the current amplitudes at voltages of -80 mV or $+80$ mV.

Full methods for all cloning, construction of expression constructs, SDS-PAGE and immunoprecipitation analyses, and construction of DT-40 cell lines are available as Supplementary Information.

Received 5 December 2000; accepted 23 April 2001.

- Harteneck, C., Plant, T. D. & Schultz, G. From worm to man: three subfamilies of TRP channels. *Trends Neurosci.* **23**, 159–166 (2000).
- Ryazanov, A. G. *et al.* Identification of a new class of protein kinases represented by eukaryotic elongation factor-2 kinase. *Proc. Natl. Acad. Sci. USA* **94**, 4884–4889 (1997).
- Romani, A. M. & Scarpa, A. Regulation of cellular magnesium. *Front. Biosci.* **5**, D720–734 (2000).
- Luthi, D., Gunzel, D. & McGuigan, J. A. Mg-ATP binding: its modification by spermine, the relevance to cytosolic Mg^{2+} buffering, changes in the intracellular ionized Mg^{2+} concentration and the estimation of Mg^{2+} by ^{31}P -NMR. *Exp. Physiol.* **84**, 231–252 (1999).

5. Runnels, L. W., Yue, L. & Clapham, D. E. TRP-PLIK, a bifunctional protein with kinase and ion channel activities. *Science* **291**, 1043–1047 (2001).

Supplementary Information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Tani and M. Monteilh-Zoller for technical assistance. This work was funded in part by a Beth Israel Pathology Foundation grant and a BIDMC Fireman Fellowship award to A.M.S., and a NIH grant. A.-L.P. is supported by the German Academic Exchange Service (DAAD).

Correspondence and requests for materials should be addressed to A.M.S. (e-mail: andrewms@u.washington.edu). The GenBank accession numbers for murine and human LTRPC7 cDNA and protein sequences are AY032951 and AY032950, respectively.

ADP-ribose gating of the calcium-permeable LTRPC2 channel revealed by Nudix motif homology

Anne-Laure Perraud^{*,†}, Andrea Fleig^{†,‡}, Christopher A. Dunn^{†,§}, Leigh Ann Bagley[‡], Pierre Launay^{*}, Carsten Schmitz^{*}, Alexander J. Stokes^{*}, Qiqin Zhu^{*}, Maurice J. Bessman[§], Reinhold Penner[‡], Jean-Pierre Kinet^{*} & Andrew M. Scharenberg^{*||}

^{*} Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA

[‡] Laboratory of Cell and Molecular Signaling, Center for Biomedical Research at The Queen's Medical Center and John A. Burns School of Medicine at the University of Hawaii, Honolulu, Hawaii 96813, USA

[§] Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA

[†] These authors contributed equally to this work.

Free ADP-ribose (ADPR), a product of NAD hydrolysis and a breakdown product of the calcium-release second messenger cyclic ADPR (cADPR), has no defined role as an intracellular signalling molecule in vertebrate systems. Here we show that a 350-amino-acid protein (designated NUDT9) and a homologous domain (NUDT9 homology domain) near the carboxy terminus of the LTRPC2/TrpC7 putative cation channel¹ both function as specific ADPR pyrophosphatases. Whole-cell and single-channel analysis of HEK-293 cells expressing LTRPC2 show that LTRPC2 functions as a calcium-permeable cation channel that is specifically gated by free ADPR. The expression of native LTRPC2 transcripts is detectable in many tissues including the U937 monocyte cell line, in which ADPR induces large cation currents (designated I_{ADPR}) that closely match those mediated by recombinant LTRPC2. These results indicate that intracellular ADPR regulates calcium entry into cells that express LTRPC2.

The LTRPC family of putative ion channel proteins² has a unique amino-terminal region of 600–700 amino acids that can be divided into 4 smaller sub-regions based on their high level of conservation in one or more family members. Next there is a region of around 300 amino acids that contains the putative pore-forming transmembrane spans, and then a region with predicted coiled-coil character, and finally a C-terminal extension of variable length and unique structure for individual LTRPC members. Figure 1a illustrates the structural elements of LTRPC2, which was originally designated TrpC7 (ref. 1), but was named LTRPC2 in a recently proposed nomenclature².

Whereas northern blotting indicates that LTRPC2 was dominantly expressed in brain (data not shown, consistent with previous reports¹), polymerase chain reaction after reverse transcription of RNA (RT-PCR) detected LTRPC2 transcripts in many other tissues, including bone marrow, spleen, heart, leukocytes, liver and lung (data not shown). Specificity of the RT-PCR analysis was confirmed by cloning a full-length LTRPC2 transcript from a human monocyte complementary DNA library. On the basis of its unique C-terminal structure, we subsequently cloned a new cDNA from a spleen cDNA library, designated NUDT9. Its expression was detectable by RT-PCR in every tissue analysed, including bone marrow, spleen, heart, leukocytes, liver, lung, kidney, prostate, testis and skeletal muscle (data not shown). The C-terminal region of LTRPC2 and NUDT9 share 50% homology and are also homologous to the *Caenorhabditis elegans* predicted protein EEED8.8 (Fig. 1b). Sequence analysis of NUDT9 revealed the presence of a putative signal peptide/anchor and a Nudix box sequence motif. Nudix boxes are found in a family of diverse enzymes that catalyse the hydrolysis of nucleoside diphosphate derivatives³. This motif is highly conserved in EEED8.8, and is present in a less conserved form in the NUDT9 homology region (NUDT9-H) of LTRPC2.

On the basis of the presence of the Nudix box in NUDT9 and the homology between NUDT9 and LTRPC2, we thought that identifying a potential substrate for NUDT9 would provide insight into LTRPC2 function. We therefore expressed NUDT9 in *Escherichia coli*, purified the protein, and screened a series of nucleoside diphosphate derivatives for enzymatic activity. The recombinant protein was a highly specific ADPR pyrophosphatase (yielding AMP and ribose 5-phosphate) with a K_m of $100 \pm 10 \mu\text{M}$ and a $V_{\max}$ of $11.8 \pm 0.3 \mu\text{mol min}^{-1} \text{mg}^{-1}$ protein. We then expressed the LTRPC2. NUDT9-H in *E. coli* and evaluated its activity towards the same panel of substrates. NUDT9-H had the same specific ADPR pyrophosphatase activity and an identical K_m ($100 \pm 10 \mu\text{M}$), but a far lower level of activity ($V_{\max} = 0.1 \mu\text{mol min}^{-1} \text{mg}^{-1}$). This may be because of the substitution of RIL and QE amino acids in LTRPC2 for the conserved REF triad and EE diad found in the Nudix motifs of NUDT9 and EEED8.8, as these are important for the catalytic activity of other Nudix hydrolases³. However, we cannot exclude that the reduced activity of the LTRPC2 NUDT9-H domain relative to that of NUDT9 may be because it is not in its native protein context.

The simplest model to relate NUDT9/NUDT9-H activity to LTRPC2 function is to have LTRPC2 as an ion channel somehow regulated by ADPR. To test this, we used a human embryonic kidney cell line (HEK-293) with tetracycline-regulated transcription of a Flag-tagged LTRPC2 construct. As can be seen in Fig. 2a, in wild-type cells, no transcript was detectable using an LTRPC2-specific probe. After tetracycline induction of cells stably transfected with a tetracycline-controlled LTRPC2 construct, substantial expression of an around 6-kilobase (kb) recombinant LTRPC2 transcript was detectable. Similarly, anti-Flag immunoreactive protein of the correct predicted molecular mass was detected in western blots only after tetracycline induction of the stably transfected cells (Fig. 2b). Finally, anti-Flag immunofluorescence indicated that a significant portion of LTRPC2 was localized at or near the plasma membrane (Fig. 2c); this led us to perform patch-clamp analyses of plasma-membrane currents. Without tetracycline induction, ADPR has no detectable effect on plasma-membrane currents (Fig. 2d). Furthermore, in the absence of ADPR in the patch pipette, basal currents in tetracycline-treated cells are the same as in wild-type HEK-293 cells. This indicates that LTRPC2 is not open constitutively under our conditions of standard intracellular solutions. In contrast, after tetracycline induction, large inward and outward currents reversing at 0 mV were induced by $100 \mu\text{M}$ ADPR (Fig. 2d; Fig. 2e, linear $I-V$ curves at various times of current development). No current activation was observed using $100 \mu\text{M}$ of a variety of closely related molecules, including NAD^+ , cADPR, ATP, ADP,

^{||} Present address: Department of Pediatrics and Immunology, University of Washington and Children's Hospital and Medical Center, Seattle, Washington 98195-6320, USA.

ADP-glucose, ADP-mannose, GDP-glucose, GDP-mannose, UDP-glucose and UDP-mannose, or ADPR's immediate breakdown products AMP and ribose 5-phosphate. Furthermore, no detectable gating occurred with agents that induce depletion of intracellular Ca^{2+} stores, such as 20 μM inositol 1,4,5-triphosphate or 10 μM ionomycin (data not shown). The relationship between intracellular ADPR concentration and whole-cell current magnitude (summarized in Fig. 2f) indicates that LTRPC2 gating begins at around 60–100 μM ADPR and saturates at around 300 μM with an effector concentration for half-maximum response (EC_{50}) of 90 μM . The EC_{50} for channel gating is similar in magnitude to the enzymatic K_m of both NUDT9 and LTRPC2 NUDT9-H, consistent with the hypothesis that this domain is directly involved in the ADPR-induced effects.

The LTRPC2 structure and linear I - V curve reversing at 0 mV both indicate that LTRPC2 functions as a non-specific cation channel. To analyse its permeation selectivity more directly, we measured ADPR-induced currents in whole cells and briefly applied solutions with isotonic substitutions for specific extracellular cations. The inward component of ADPR-induced whole-cell currents was suppressed when cells were perfused extracellularly with isotonic N -methyl-D-glucamine chloride (Fig. 3a) or choline-chloride (not shown). Similar experiments, using isotonic Ca^{2+} solutions (120 mM CaCl_2) revealed the ability of Ca^{2+} ions to maintain approximately 50% of the inward current observed using the standard extracellular solution (Fig. 3b). These data indicate that LTRPC2 functions as a non-selective cation channel that is permeant to both Na^+ and Ca^{2+} .

To address the question of whether ADPR directly affects LTRPC2 function, we tested the ability of ADPR to induce LTRPC2-gating in cell-free, inside-out patches. As shown in Fig. 3c, under the same conditions used for the whole-cell analyses, channel openings were easily detectable, within seconds of brief applications of ADPR to

the inside surface of patches excised from induced HEK-293 cells, but not from patches excised from uninduced cells. Channels opened with concentrations of ADPR as low as 10 μM , albeit in only one out of five patches. At concentrations of 30 μM or higher, ADPR consistently gated multiple channels in all patches ($n = 12$). Remarkably, brief applications of ADPR resulted in sustained channel openings for periods of up to 1–2 min even after ADPR exposure was stopped; this may in part contribute to the high apparent cooperativity of LTRPC2's dose-response curve for ADPR. Channel activity could be induced repetitively and reversibly in the same patch for up to 20 min with no obvious signs of decreased activity. We also analysed the channel activity using symmetric NaCl-based solutions on each side of the patch mem-

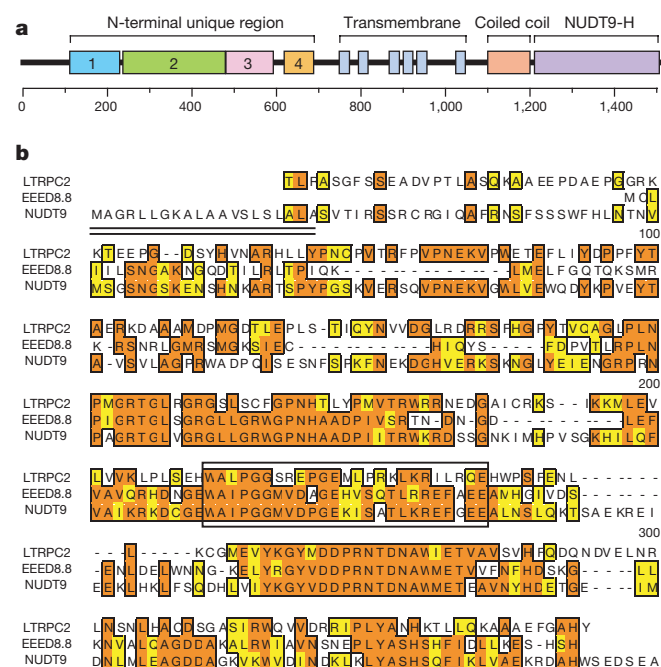


Figure 1 Protein sequence analysis of LTRPC2. **a**, Schematic of LTRPC2 structural motifs on the basis of alignments of various related proteins including MLSN-1, LTRPC7, MTR-1 and the *C. elegans* proteins C05C12.3, T01H8.5 and F54D1.5. **b**, ClustalW alignment of the NUDT9 homology region of LTRPC2, EEED8.8 and NUDT9. The putative signal peptide or anchor found in NUDT9 is double underlined (prediction on the basis of SignalP2.0 analysis of the NUDT9 amino-acid sequence). The Nudix box region is boxed in a rectangle.

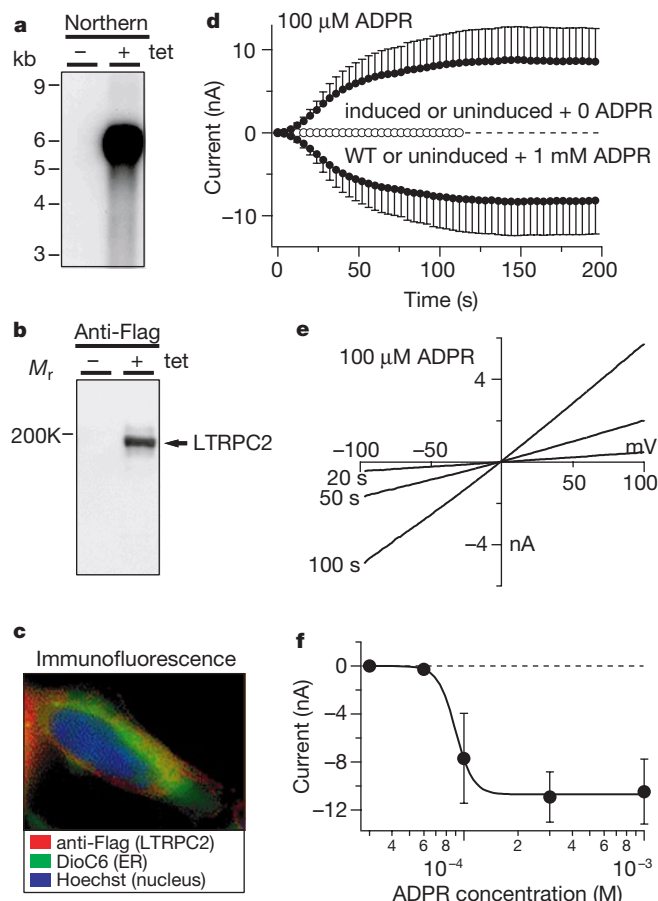


Figure 2 Functional expression of LTRPC2 in HEK-293 cells. **a**, Non-transfected wild-type or tetracycline-induced HEK-293 cells ($1 \mu\text{g ml}^{-1}$, 24 h) were analysed by northern blot using a human LTRPC2 probe. Recombinant LTRPC2 is revealed as an ~5.5-kb messenger RNA. **b**, HEK-293 cells were treated or not for 24 h with $1 \mu\text{g ml}^{-1}$ tetracycline. Cells (10^6) were analysed for expression of a Flag-reactive protein by anti-Flag immunoprecipitation/anti-Flag immunoblotting. **c**, Representative HEK-293 cell after tetracycline induction of Flag-LTRPC2 expression and staining with anti-Flag monoclonal antibody (red), DioC6 (green, perinuclear endoplasmic reticulum) and Hoechst (blue, nucleus). Peripheral red staining indicates LTRPC2 in the plasma membrane. **d**, Temporal development of averaged membrane currents at -80 and +80 mV. Only tetracycline-induced HEK-293 cells expressing Flag-LTRPC2 generated large inward and outward currents when perfused with 100 μM ADPR ($n = 4 \pm \text{s.e.m.}$, filled symbols). Open symbols represent superimposed analyses of responses from wild-type or uninduced HEK-293 cells perfused with or without 1 mM ADPR ($n = 3-5$). **e**, I - V relationships of ADPR-induced currents at different times after break in a representative cell perfused with 100 μM ADPR. **f**, Dose-response curve for ADPR-dependent gating of LTRPC2 on the basis of average currents at -80 mV ($n = 5-12 \pm \text{s.e.m.}$), yielding an apparent EC_{50} of 90 μM and a Hill coefficient of 9 (Hill coefficients of 4–8 yielded similarly adequate approximations).

brane (Fig. 3d), with $[Ca^{2+}]_i$ buffered to 100 nM and with the solution facing the inside patch surface supplemented with 100 μ M ADPR. Again, single-channel events were readily detected upon patch excision into this solution ($n = 8$); however, under these conditions, the probability of channel opening was voltage dependent, with shorter channel opening times observed at negative potentials. The single-channel current–voltage relationship is shown in Fig. 3e and linear regression through the open levels of single-channel events yielded a slope conductance of 60 pS. Because the voltage dependence exhibited by single channels under these symmetrical Na^+ -based conditions had not been observed in our previous whole-cell analyses, we measured whole-cell currents using these ionic conditions. We observed a corresponding voltage-

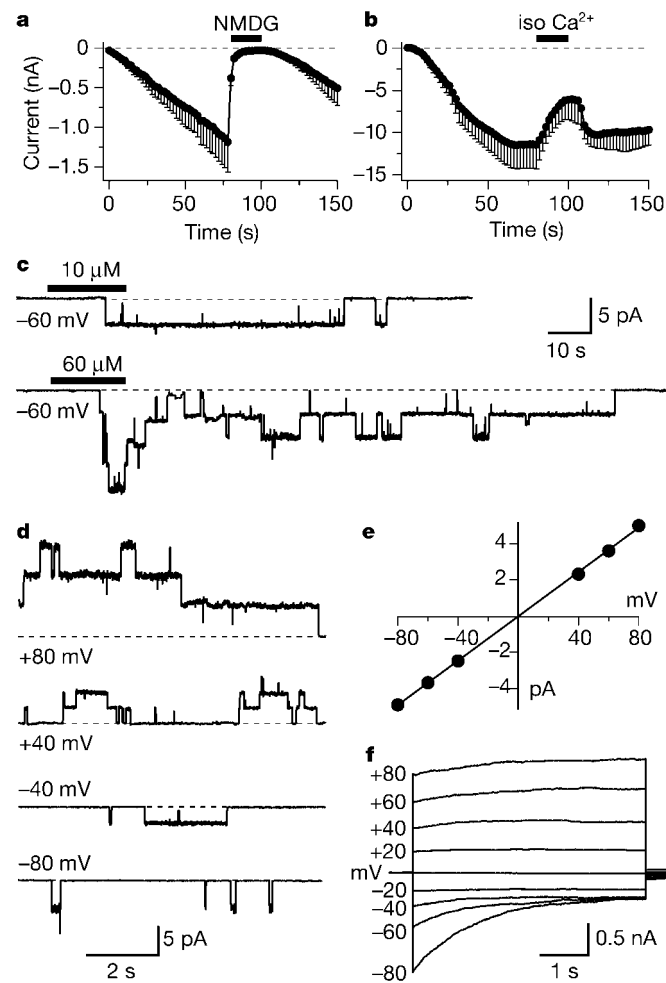


Figure 3 Permeation and single-channel properties of LTRPC2. **a**, HEK-293 cells expressing Flag-LTRPC2 were perfused with 300 μ M ADPR. Experiments were performed on cells after 16 h induction, resulting in smaller average current amplitudes. At the time indicated by the bar, isotonic NMDG-Cl solution (180 mM NMDG-Cl, 330 mOsm) was applied externally ($n = 3$). **b**, Same as **a**, except that isotonic $CaCl_2$ solution (120 mM $CaCl_2$, 300 mOsm) was applied ($n = 3$), supporting about 50% of current previously carried by Na^+ . **c**, Reversible activation/deactivation of LTRPC2 by 10 μ M and 60 μ M ADPR, briefly applied to inside-out patches using Cs-glutamate-based internal and NaCl-based external solutions. Channel open times were extremely long under these recording conditions. **d**, LTRPC2 channels recorded under symmetrical Na^+ -based solutions with 100 μ M ADPR added to the internal solution. There was an apparent voltage dependence of channel opening probability. **e**, Single-channel I - V relationship derived from averages of several events from different patches ($n = 3$ –12), yielding a single-channel conductance of 60 pS. **f**, Average whole-cell currents using the same solutions as in **d**. At negative potentials the channels inactivated to a steady-state plateau, whereas positive potentials recruited more channels over time ($n = 6$).

dependent deactivation of whole-cell currents on a seconds time scale at negative potentials (Fig. 3f). This shows that the behaviour of macroscopic ADPR-induced currents mirrors that observed for ADPR-gated single channels. These results provide compelling evidence that ADPR directly gates LTRPC2.

The data indicate that free ADPR may be able to regulate calcium entry by directly gating LTRPC2, and they also provide several elements that may be used to identify native calcium-entry pathways that use this mechanism. However, there are neither published reports linking ADPR to calcium entry nor any reports of native ADPR-induced membrane currents in mammalian systems. As the source of our LTRPC2 cDNA clone was a cDNA library of human monocytes, we evaluated the human U937 monocyte cell line for expression of LTRPC2 transcripts by northern blotting. As illustrated in Fig. 4a, we detected a transcript in U937 cells that was slightly larger than that of our inducible LTRPC2-expressing HEK-293 cell line, and this was consistent with the predicted 6.2 kb size of the native LTRPC2.

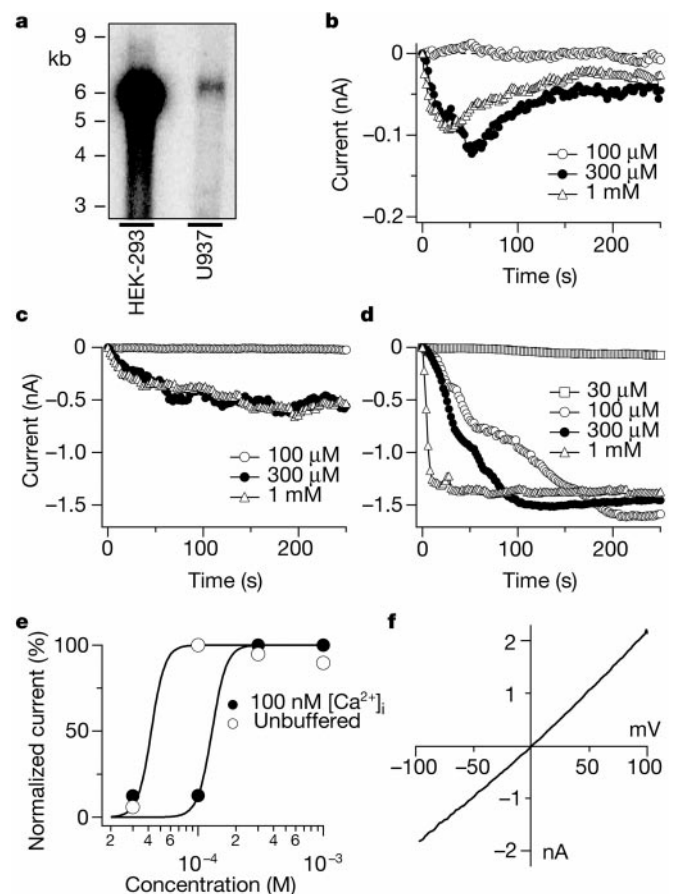


Figure 4 ADPR-gated cation currents in human U937 cells. **a**, Northern blot analysis identifies LTRPC2 as a 6 kb mRNA in tetracycline-induced HEK-293 cells (1μ g ml^{-1} , 24 h) and in native U937 cells. This blot was exposed for a longer time to provide optimal detection of the native transcript, causing overexposure of the recombinant transcript. **b**, Inward currents in U937 cells at -80 mV activated by different intracellular concentrations of ADPR in the presence of 10 mM BAPTA ($n = 4$ –11 each). **c**, Same as **b**, except that $[Ca^{2+}]_i$ was buffered to 100 nM ($n = 5$ –7 each). **d**, Same as **b** and **c** in the absence of exogenous buffers ($n = 5$ –9 each). **e**, Dose–response relationships for I_{ADPR} in U937 cells perfused with defined ADPR concentrations while $[Ca^{2+}]_i$ was buffered to 100 nM (filled symbols) or left to vary freely by omitting exogenous buffers (open symbols), yielding an apparent EC_{50} of 130 μ M and 40 μ M for buffered and unbuffered conditions, respectively (in both cases, Hill coefficients were 8). **f**, I - V relationship of ADPR currents in U937 cells. Representative current records in response to a voltage ramp ranging from -100 to $+100$ mV over 50 ms, obtained with 100 μ M ADPR under unbuffered conditions.

We performed whole-cell patch-clamp analyses of U937 cells with varying amounts of ADPR present in the intracellular solution (Fig. 4b). Once the whole-cell configuration was established, inward currents developed rapidly if ADPR was present in the patch pipette. Under these experimental conditions, in which 10 mM BAPTA was employed to buffer $[Ca^{2+}]_i$ to zero, the EC_{50} of ADPR was around 130 μ M, comparable with the amount required for gating of LTRPC2 in the recombinant system (90 μ M). The currents produced by ADPR (designated I_{ADPR}) were small with an inactivation that was time-dependent. This inactivation was not observed, and current amplitudes were significantly larger, when free $[Ca^{2+}]_i$ was buffered to 100 nM with BAPTA (Fig. 4c). This indicates that I_{ADPR} may possess some degree of Ca^{2+} -dependence with a decreased probability of channel opening when BAPTA-containing pipette solutions are equilibrated with the cytosol and clamp $[Ca^{2+}]_i$ to zero. This interpretation is further supported by experiments in which $[Ca^{2+}]_i$ was left to vary freely (that is, omitting BAPTA and leaving cytosolic Ca^{2+} concentrations unbuffered). These conditions led to an I_{ADPR} developing more rapidly and of an even larger magnitude (Fig. 4d). Dose-response curves for ADPR revealed a similar steepness to that observed for recombinant LTRPC2 channels and, under unbuffered conditions, there was a significant leftward shift of the dose-response curve to ADPR (yielding an EC_{50} of 40 μ M; Fig. 4e). We observed a similar sensitization in the LTRPC2-expressing HEK-293 cells by $[Ca^{2+}]_i$ (data not shown) and noticed a larger degree of variability in responses in both HEK-293 and U937 cell systems, as would be expected when $[Ca^{2+}]_i$ was left to vary freely. Regardless of the above experimental conditions, I_{ADPR} had a linear I - V relationship reversing at 0 mV that was indistinguishable from that observed with currents carried by heterologously expressed LTRPC2 (Figs 4f and 2f).

In summary, our results provide converging biochemical and electrophysiological evidence that ADPR can regulate calcium entry in vertebrate systems through its ability to gate LTRPC2 directly. They also provide significant insight into LTRPC2 structure/function by raising the possibility that the NUDT9-H domain of LTRPC2 can function as an ADPR binding site and/or act as an intrinsic deactivation mechanism for LTRPC2 through its ability to break down ADPR. Our results provide an unequivocal molecular link between ADP-ribose and calcium entry into a mammalian system. It has been reported that ADPR gates the fertilization channel of ascidian oocytes⁴, but differences in single-channel conductance as well as the differential effects of intracellular Ca^{2+} buffering seem to set these two ion channels apart. As free ADPR is produced in many contexts, for example, through the pyridine nucleotide cycle, apoptosis, cytosolic and ecto-NAD glycohydrolases, cADPR breakdown and within mitochondria^{5–12}, ADPR-mediated gating of LTRPC2 may have numerous roles in cellular function. □

Methods

RT-PCR and northern blot analysis of expression

For PCR analysis of LTRPC2 expression, we used the primers CAGTGTGGCTACAGC-CATGA and TCAGGCCCGTGAAGACGATG to obtain a 138-base pair (bp) band. For NUDT9 expression, we used the primers GGCAAGACTATAAGCCTGTG and ATAATGGGATCTGCAGCGTG to obtain a 252-bp band. All libraries screened were from Life Technologies. For northern blots, single-stranded probes were constructed with the *NotI*/*Bgl*III fragment of the human LTRPC2 sequence as template. Hybridizations were performed using standard methods.

Assays for Nudix activity

Enzyme velocities were quantified by measuring the conversion of a phosphatase-insensitive substrate, ADPR, to the phosphatase-sensitive products, AMP and ribose-5-phosphate. The standard incubation mixture (50 μ l) contained 50 mM Tris-Cl buffer (pH 9.0), 16 mM $MgCl_2$, 2 mM ADPR, 0.2–1 milliunits of enzyme and 4 units of alkaline intestinal phosphatase. After 30 min at 37 °C, the reaction was stopped by adding EDTA buffer and inorganic orthophosphate was measured as described¹³. A unit of enzyme hydrolysed 1 μ mol of substrate per min under these conditions, and 2 moles of phosphate were liberated per mole of ADPR hydrolysed. The Nudix-type substrates tested included

ATP/deoxy-ATP, CTP/deoxy-CTP, GTP/deoxy-GTP, TTP/deoxy-TTP, GDP-mannose, ADP-glucose, UDP-glucose, Ap_nA ($n = 2–6$) (diadenosine polyphosphates, where the number of phosphates are 2–6), NADH, NAD^+ , cADPR and ADPR.

The standard assay mixture (without alkaline intestinal phosphatase) was incubated for 30 min at 37 °C and stopped by adding 50 μ l of a mixture of four parts of Norit (20% packed volume) and one part of 7% $HClO_4$ to remove adenine-containing nucleotides. After centrifugation, 50 μ l was adjusted to an alkaline pH and incubated for an additional 30 min at 37 °C with alkaline intestinal phosphatase to hydrolyse the ribose-5-phosphate formed. The subsequent free phosphate was measured and compared with a control reaction that was not treated with Norit. The stoichiometric relation between the two indicated that the products were AMP and ribose-5-phosphate.

SDS-PAGE and immunoassays

SDS-PAGE, immunoprecipitation and immunoblotting were performed using standard methods or as described in the Figure legends. For immunofluorescence, after 24 h tetracycline induction, HEK-293 cells were fixed (4% paraformaldehyde, 20 min) and permeabilized (0.2% triton X-100, 4 min) before sequential exposure to Hoechst (1 μ g ml^{-1} , 2 min) and DioC6 (0.3 μ g ml^{-1} , 2 min) (Molecular Probes). For anti-Flag immunofluorescence, cells were blocked and probed with anti-Flag antibody (IBI-Kodak), and then with Alexa 568 goat anti-mouse IgG (Molecular Probes). Mounted samples were imaged using single emission filters (Texas Red, FITC and Hoechst).

Cell culture and electrophysiology

Wild-type and tetracycline-inducible HEK-293 Flag-LTRPC2-expressing cells were cultured at 37 °C with 5% CO_2 in DMEM supplemented with 10% fetal bovine serum and 2 mM glutamine. The medium was supplemented with blasticidin (5 μ g ml^{-1} ; Invitrogen) and zeocin (0.4 mg ml^{-1} ; Invitrogen). Cells were resuspended in medium containing 1 μ g ml^{-1} tetracycline (Invitrogen) 24 h before experiments. For patch-clamp experiments, cells were kept in standard Ringer's solution (in mM): 145 NaCl, 2.8 KCl, 1 $CaCl_2$, 2 $MgCl_2$, 10 glucose, 10 HEPES-NaOH (pH 7.2). Standard pipette-filling solutions contained (in mM): 145 Cs-glutamate, 8 NaCl, 1 $MgCl_2$, 10 Cs-BAPTA; 10 HEPES-CsOH (pH 7.2 adjusted with CsOH). In some experiments, $[Ca^{2+}]_i$ was buffered to 100 nM with 10 mM BAPTA and 3.6 mM $CaCl_2$ or left unbuffered. All nucleotides and reagents were purchased from Sigma and dissolved in the standard intracellular solution. Patch-clamp experiments were performed in the whole-cell configuration at 21–25 °C. Data were acquired with 'Pulse' software controlling an EPC-9 amplifier (HEKA). Voltage ramps of 50 ms spanning the voltage range from –100 to +100 mV were delivered from a holding potential of 0 mV at a rate of 0.5 Hz over a period of 200–400 s. When applicable, voltages were corrected for liquid junction potentials. Currents were filtered at 2.9 kHz and digitized at 100 μ s intervals. Capacitive currents and series resistance were determined and corrected before each voltage ramp. For analysis, the very first ramps before activating the currents were digitally filtered at 2 kHz, pooled and used for leak-subtraction of all subsequent current records. The low-resolution temporal development of currents for a given potential was extracted from the leak-corrected individual ramp current records by measuring the current amplitudes at voltages of –80 mV or +80 mV. Single channels were recorded in the inside-out configuration and currents were filtered and sampled as above. For display purposes, data records were digitally filtered and down-sampled to 100 Hz.

For details of methods used for LTRPC2 and NUDT9 cloning, construction of eukaryotic and prokaryotic expression vector and NUDT9/NUDT9-H purification, see Supplementary Information.

Received 5 December 2000; accepted 16 March 2001.

- Nagamine, K. *et al.* Molecular cloning of a novel putative Ca^{2+} channel protein (TRPC7) highly expressed in brain. *Genomics* **54**, 124–131 (1998).
- Harteneck, C., Plant, T. D. & Schultz, G. From worm to man: three subfamilies of TRP channels. *Trends Neurosci.* **23**, 159–166 (2000).
- Bessman, M. J., Frick, D. N. & O'Handley, S. F. The MuT proteins or 'Nudix' hydrolases, a family of versatile, widely distributed, 'housecleaning' enzymes. *J. Biol. Chem.* **271**, 25059–25062 (1996).
- Wilding, M., Russo, G. L., Galione, A., Marino, M. & Dale, B. ADP-ribose gates the fertilization channel in ascidian oocytes. *Am. J. Physiol.* **275**, C1277–1283 (1998).
- Hillyard, D. *et al.* The pyridine nucleotide cycle. Studies in *Escherichia coli* and the human cell line D98/AH2. *J. Biol. Chem.* **256**, 8491–8497 (1981).
- Dousa, T. P., Chini, E. N. & Beers, K. W. Adenine nucleotide diphosphates: emerging second messengers acting via intracellular Ca^{2+} release. *Am. J. Physiol.* **271**, C1007–1024 (1996).
- McConkey, D. J. & Orrenius, S. Signal transduction pathways in apoptosis. *Stem Cells* **14**, 619–631 (1996).
- Koch-Nolte, F. & Haag, F. Mono(ADP-ribosyl)transferases and related enzymes in animal tissues. Emerging gene families. *Adv. Exp. Med. Biol.* **419**, 1–13 (1997).
- Okazaki, I. J. & Moss, J. Glycosylphosphatidylinositol-anchored and secretory isoforms of mono-ADP-ribosyltransferases. *J. Biol. Chem.* **273**, 23617–23620 (1998).
- Chakraborti, T., Das, S., Mondal, M., Roychoudhury, S. & Chakraborti, S. Oxidant, mitochondria and calcium: an overview. *Cell. Signal.* **11**, 77–85 (1999).
- Liang, M., Chini, E. N., Cheng, J. & Dousa, T. P. Synthesis of NAADP and cADPR in mitochondria. *Arch. Biochem. Biophys.* **371**, 317–325 (1999).
- Guse, A. H. Cyclic ADP-ribose. *J. Mol. Med.* **78**, 26–35 (2000).
- Ames, B. N. & Dubin, D. T. The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. *J. Biol. Chem.* **235**, 769–775 (1960).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Tani and M. Monteilh-Zoller for technical assistance. This work was funded in part by a Beth Israel Pathology Foundation grant, a BIDMC Fireman Fellowship award to A.M.S. and NIH grants to A.M.S. and M.J.B. A.L.P. is supported by the German Academic Exchange Service (DAAD). In memory of Q. Zhu.

Correspondence and requests for materials should be addressed to A.M.S. (e-mail: andrewms@u.washington.edu). The GenBank accession number for NUDT9 cDNA and protein sequences is AY026252.

Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

Jean-Pierre Hugot^{*,†,‡,§,¶,||,¶¶}, Mathias Chamaillard^{*,†,‡}, Habib Zouali^{*,§}, Suzanne Lesage^{*,†,‡}, Jean-Pierre Cézard[‡], Jacques Belaiche[§], Sven Almer^{||}, Curt Tysk[¶], Colm A. O'Morain[#], Miquel Gassull[☆], Vibeke Binder^{*,†}, Yigael Finkel^{†,‡}, Antoine Cortot^{‡,§}, Robert Modigliani^{§,§}, Pierre Laurent-Puig[†], Corine Gower-Rousseau^{‡,§}, Jeanne Macry^{||}, Jean-Frédéric Colombel^{‡,§}, Mourad Sahbatou^{*,†,¶} & Gilles Thomas^{*,†,¶}

* Fondation Jean Dausset CEPH, 27 rue J. Dodu 75010 Paris, France

† INSERM U434, 27 rue J. Dodu 75010 Paris, France

‡ PEWG-IBD, Department of Paediatric Gastroenterology, Hôpital Robert Debré, 75019 Paris, France

§ GETAID and Department of Gastroenterology, CHU de Liège, 4000 Belgium

|| Division of Gastroenterology and Hepatology, IHM, Linköping Universitet, SE-581 85 Linköping, Sweden

¶ Department of Gastroenterology, Örebro Medical Center Hospital, SE-701 85 Örebro, Sweden

Department of Gastroenterology, Adelaide & Meath Hospital, Dublin 24, Ireland

☆ Department of Gastroenterology, Hospital Universitari Germans Trias i Pujol, Badalona, 08916 Spain

** Department of Gastroenterology Herlev Hospital, 2730 Herlev, Denmark

†† Astrid Lindgren Children's Hospital, SE-161 76 Stockholm, Sweden

‡‡ Registre EPIMAD, Hôpital Calmette, 59037 Lille, France

§§ Department of Gastroenterology, Hôpital Saint Louis, 75475 Paris, France

||| INSERM U458, Hôpital Robert Debré, 75019 Paris, France

¶¶ Department of Surgery, Hôpital Saint Antoine, 75012 Paris, France

Crohn's disease^{1,2} and ulcerative colitis, the two main types of chronic inflammatory bowel disease, are multifactorial conditions of unknown aetiology. A susceptibility locus for Crohn's disease has been mapped³ to chromosome 16. Here we have used a positional-cloning strategy, based on linkage analysis followed by linkage disequilibrium mapping, to identify three independent associations for Crohn's disease: a frameshift variant and two missense variants of *NOD2*, encoding a member of the Apaf-1/Ced-4 superfamily of apoptosis regulators that is expressed in monocytes. These *NOD2* variants alter the structure of either the leucine-rich repeat domain of the protein or the adjacent region. *NOD2* activates nuclear factor NF- κ B; this activating function is regulated by the carboxy-terminal leucine-rich repeat domain, which has an inhibitory role and also acts as an intracellular receptor for components of microbial pathogens. These observations suggest that the *NOD2* gene product confers susceptibility to Crohn's disease by altering the recognition of these components and/or by over-activating NF- κ B in monocytes, thus documenting a molecular model for the pathogenic mechanism of Crohn's disease that can now be further investigated.

Crohn's disease (CD; MIM 266600) occurs primarily in young

adults, with an estimated prevalence of 1 in 1,000 in western countries¹. Its incidence has increased markedly over the past half century, arguing for the involvement of recent, unidentified, environmental factors². Familial aggregation of the disease suggests that genetic factors may also be involved—an hypothesis that was substantiated in 1996 by the discovery of a susceptibility locus for CD, *IBD1*, on chromosome 16 (ref. 3). Identification of the exact nature of the genetic changes that are implicated in CD susceptibility would provide a specific approach to understanding this common disorder.

Because candidate genes previously localized on chromosome 16 failed to show an association with CD^{4,5}, we refined the localization of the *IBD1* susceptibility locus by typing 26 microsatellite markers spaced at an average distance of 1 cM in the pericentromeric region

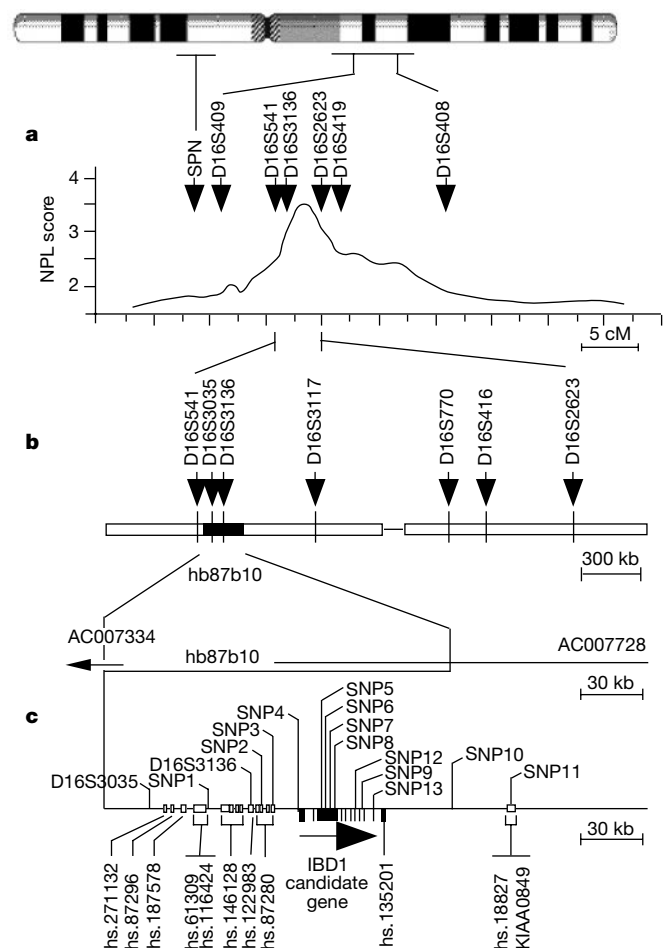


Figure 1 Strategy used to identify the *IBD1* locus. **a**, Profile for the multipoint non-parametric linkage (NPL) scores. Approximate cytogenetic localizations are shown for selected microsatellite markers used in the study³ that localized *IBD1* to the pericentromeric region of chromosome 16. Subsequent linkage analyses were focused on the region between SPN and D16S408. The highest NPL score (maximum NPL score, 3.49; $P = 2.37 \times 10^{-4}$) was in the region between markers D16S541 and D16S2623 (ref. 6). **b**, Physical map of the *IBD1* region⁶. White and black boxes correspond to the two BAC contigs and BAC clone hb87b10, respectively. Five yeast artificial chromosomes (YACs) bridge a gap of ~100 kb between these two contigs. The position on the physical map of the microsatellite markers used in the linkage analysis is indicated. Distance between D16S541 and D16S2623 is ~2 Mb. **c**, Representation of the sequenced region containing the *IBD1* candidate gene. Unigene clusters and 11 exons of the *IBD1* candidate gene are indicated by white and black boxes, respectively. Bold horizontal arrow indicates direction of transcription. Positions of SNP 1–13, D16S3035 and D16S3136, which were typed in 235 CD families for linkage disequilibrium studies, are shown.

of chromosome 16. Model-free linkage analyses performed on 77 multiplex families indicated that the probability was 0.7 for the location of the susceptibility locus between D16S541 and D16S2623 (Fig. 1a). We constructed bacterial artificial chromosome (BAC) contigs spanning this region (Fig. 1b), which supported linkage disequilibrium mapping⁶. The transmission disequilibrium test⁷, performed on a single trio from each of 108 (77 multiplex and 31 simplex) families, showed a borderline significant association ($P < 0.05$) between the disease phenotype and the 207-base-pair (bp) allele of D16S3136. This observation was replicated with another set of 76 families, although with a different allele (the 205-bp allele; $P < 0.01$). These two observations may be due to type-one errors. Alternatively, they may reflect true association in two sets of families drawn from genetically different populations.

The latter hypothesis led to the following strategy: a 164-kb BAC clone (hb87b10) from the CEPH-BAC library containing D16S3136 was sequenced completely (EMBL accession number AJ303140). A public database search extended the sequence of the corresponding region to 260 kb but did not identify characterized genes, with the exception of KIAA0849, which codes for a ubiquitin C-terminal hydrolase homologue in *Caenorhabditis elegans*. However, analysis by GRAIL and an expressed sequence tag (EST) homology search identified many putatively transcribed regions (Fig. 1c).

Eleven single-nucleotide polymorphisms (SNP 1–11) selected from these regions were genotyped, together with microsatellite markers D16S3035 and D16S3136, in a total of 235 available CD families (Table 1). Strong linkage disequilibrium was observed among most markers (data not shown). Several SNPs showed significant association with CD by the pedigree disequilibrium test⁸ (PDT), confirming the existence of linkage disequilibrium, with the disease locus over the investigated region (specially SNP 2, nominal P value 0.00002; Table 1).

These observations prompted the characterization of neighbouring Unigene clusters (Fig. 1c). Eleven overlapping clones, isolated

from a human leukocyte complementary DNA library, extended Unigene cluster hs135201 and identified 11 exons of a single gene. The previously identified SNP 5–8 were contained in exon 3 of this gene and shown to be non-synonymous variants. To find additional disease-related variants, all exons of this gene were sequenced in 50 unrelated CD patients—each a member of an affected sibling pair identical by descent for both chromosome 16 homologous regions. Two additional non-synonymous SNPs (SNP 12 and 13), with rare-allele frequencies greater than 0.03, were identified and subsequently used to type the 235 CD families.

The PDT was most significant for SNP 13 ($P = 6 \times 10^{-6}$). Families were divided into two groups: those with at least one member carrying the rare allele of SNP 13 and those without this allele. The latter group of families failed to show association between CD and SNP 4–6, and showed considerable decrease in the significance of the SNP 2 association. This result indicates that the associations of these four loci with CD were not independent of SNP 13 (Table 1). In contrast, significance of the CD associations with SNP 8 and 12 decreased modestly in these families, indicating a minimal contribution of the rare SNP 13 allele to these associations.

The 8 intragenic SNPs that were initially identified defined 41 different haplotypes. Three of these revealed preferential transmission to affected individuals (Table 2). These three haplotypes each contain one rare allele of SNP 8, 12 or 13 in a context of a common background. Notably, the haplotype defined by the same background and by the absence of these rare alleles did not show such transmission distortion (Table 2). Furthermore, the rare alleles of SNP 8, 12 and 13 were never found on the same haplotype, indicating independent association of CD susceptibility with each of three non-synonymous variants of a same gene.

As a result of these associations, the allele frequencies of SNP 8, 12 and 13 differed in the group of CD patients as compared with controls (Table 3). Average risks for CD, computed for genotypes containing zero, one or two variants (Table 3), revealed a gene-

Table 1 Linkage disequilibrium analyses at the *IBD1* locus

Marker name	Distance to next marker (bp)	Rare allele frequency in CD family founders*	Nominal P -values for the PDT in CD families†			Marker location/structural features‡
			All ($N = 235$)	SNP 13+ ($N = 65$)	SNP 13– ($N = 170$)	
D16S3035	29,389	Microsatellite marker 0.34	NS	0.005	NS	End of BAC hb133D1
SNP1	19,434		NS	NS	NS	
D16S3136	3,273	Microsatellite marker	NS	0.002	NS	Exon of hs.122983
SNP2	5,459		0.00002	0.00006	0.02	Exon of hs.87280
SNP3	15,076		NS	0.005	NS	Exon of hs.87280
SNP4	14,396		0.0006	0.000003	NS	Grail putative exon
SNP5	576		0.0001	0.00001	NS	<i>IBD1</i> exon 3 721C > T P241S
SNP6	385	0.38	0.0001	0.000006	NS	<i>IBD1</i> exon 3 1296C > T R432R
SNP7	344	0.35	NS	0.0003	NS	<i>IBD1</i> exon 3 1680T > T R432R
SNP8	10,593	0.10	0.001	0.03	0.009	<i>IBD1</i> exon 3 2023C > T R675W
SNP12	2,727	0.03	0.003	NS	0.0041	<i>IBD1</i> exon 7 2641G > C G1881R
SNP9	4,510	0.37	0.01	0.00003	NS	<i>IBD1</i> intron 8 IVS8-133delAinsCT
SNP13	35,121	0.07	0.000006	0.000006	–	<i>IBD1</i> exon 10 2936insC 980fs981X (frame shift)
SNP10	28,592	0.37	0.05	0.002	NS	End of BAC hb27G11
SNP11		0.09	NS	NS	NS	Exon 9 of KIAA0849 1056C > T D352D

Polymorphic markers are ordered from centromere to telomere. The nomenclature used does not take into account an alternative NOD2 exon located upstream of *IBD1* exon 1 (ref. 14). NS, not significant.
* Rare allele frequencies computed for the 589 CD family founders.
† Nominal P values for pedigree disequilibrium test (PDT) are reported for all families and the subgroups of families with or without the rare allele of SNP 13.
‡ SNP location and, when relevant, its effect on the coding sequence are indicated.

dosage effect. The major increase in risk associated with two variant alleles confirms the recessive nature of CD susceptibility, which was suggested by previous segregation analysis^{9–11} and linkage studies³. It may also contribute to explain the unusual precision of the affected sibling-pair analysis in mapping the susceptibility gene.

The rare allele of SNP 8 was associated positively with the 205-bp allele of D16S3136 and negatively with the 207-bp allele. The inverse association was noted for the rare alleles of SNP 12 and 13, thus providing a rationale for the initial observations made with this microsatellite marker (data not shown). Genotype frequencies were comparable in CD patients originating from uniquely and multiply affected kindred—an observation compatible with the close clinical similarity of the sporadic and familial diseases¹². The observed linkage of CD to chromosome 16 could not be entirely explained by the present associations, because GeneHunter analysis of 85 multiplex families without SNP 8, 12 and 13 revealed a component of linkage (nonparametric lod score (NPL) 1.6, pointwise significance $P < 0.02$). Thus, other variants of this gene or additional genes on chromosome 16 may be involved in CD susceptibility.

Genotyping of 167 patients with ulcerative colitis revealed genotype frequencies comparable to those of controls, indicating that these SNPs were not associated with susceptibility to ulcerative colitis—an observation in agreement with its lack of linkage to the *IBD1* locus¹³.

The candidate *IBD1* gene has high expression in leukocytes, but low or no expression in the other investigated tissues, including

normal colon and small intestine (results not shown). It encodes a 1,013-amino-acid protein that is identical to NOD2—a member of the CED4/APAF1 superfamily of apoptosis regulators¹⁴. From its amino terminus to its carboxy terminus, NOD2 is composed of two caspase recruitment domains (CARD), a nucleotide-binding domain (NBD) and a LRR region (Fig. 2b). The LRR domain of NOD2 has binding activity for bacterial lipopolysaccharides¹⁵ (LPS) and its deletion stimulates the NF- κ B pathway^{16–18}.

The rare allele of SNP 13 corresponds to a 1-bp insertion in exon 10 (980fs) predicted to truncate NOD2 in the LRR region. Those of SNP 8 and 12 cause non-conservative substitutions in the LRR domain (G881R) and in the proximal adjacent region (R675W), respectively (Fig. 2a). Systematic sequencing of the coding sequence of NOD2 revealed additional very rare missense variants, which together were observed in 5% of controls and 4% of patients with ulcerative colitis. This percentage rose to 17% for CD patients, where the most frequent variants tended to cluster in the LRR and its adjacent regions (Fig. 2b). This excess suggests that, in addition to SNP 8, 12 and 13, more variants in this part of the NOD2 protein may be associated with CD susceptibility. Thus, the LRR domain of CD-associated variants is likely to be impaired, possibly to various degrees, in its recognition of microbial components and/or in the physiological inhibition of NOD2 dimerization, thus resulting in the inappropriate activation of NF- κ B in monocytes.

Much evidence supports bacteria-induced NF- κ B dysregulation in CD. First, susceptibility to spontaneous inflammatory bowel

Table 2 Transmission disequilibrium of *NOD2* haplotypes in CD families

SNP no.	4	5	6	Haplotype		12	9	13	Transmitted	Non-transmitted	PDT (<i>P</i> -value)
				7	8						
SNP 8	1	1	1	1	2	1	1	1	4	1	0.001
	2	1	1	1	2	1	1	1	2	0	
	2	1	1	1	2	1	2	1	82	48	
	1	1	1	1	2	1	2	1	3	4	
	2	1	2	1	2	1	2	1	7	0	
	2	2	2	1	2	1	2	1	0	1	
SNP 12	1	2	2	2	1	2	1	1	1	3	0.0005
	2	1	2	1	1	2	1	1	0	1	
	2	1	1	1	1	2	2	1	38	13	
	1	1	1	1	1	2	2	1	3	2	
	2	2	1	1	1	2	2	1	1	1	
	2	2	2	1	1	2	2	1	2	2	
	2	2	2	2	1	2	2	1	3	1	
	1	2	2	2	1	2	2	1	1	0	
SNP 13	1	1	1	2	1	2	2	1	1	0	0.0000002
	2	1	1	1	1	1	1	2	2	1	
	2	1	1	1	1	1	2	2	83	22	
	1	1	1	1	1	1	2	2	2	0	
	1	2	2	1	1	1	2	2	1	0	
	2	2	2	2	1	1	2	2	0	1	
None*	1	2	2	2	1	1	1	1	116	141	NS
	2	2	2	2	1	1	1	1	2	4	
	1	1	2	2	1	1	1	1	0	0	
	1	2	1	2	1	1	1	1	0	2	
	1	1	1	2	1	1	1	1	1	0	NS
	2	1	1	2	1	1	1	1	1	1	
	1	2	2	1	1	1	1	1	9	19	
	1	1	1	1	1	1	1	1	4	7	
	2	1	1	1	1	1	1	1	0	0	
	2	2	2	1	1	1	1	1	2	0	
	2	2	2	2	1	1	2	1	7	4	
	1	2	2	2	1	1	2	1	9	7	
	2	1	2	1	1	1	2	1	0	1	
	1	1	2	1	1	1	2	1	1	1	
	1	2	2	1	1	1	2	1	94	116	NS
	2	2	2	1	1	1	2	1	20	16	
	2	1	1	1	1	1	2	1	70	78	
	1	1	1	1	1	1	2	1	11	7	
	2	2	1	1	1	1	2	1	0	1	NS
	1	2	1	1	1	1	2	1	3	9	

All haplotypes defined by typing 8 intragenic SNPs of *IBD1* in 235 CD families are indicated. They are classified according to the presence of a rare allele of SNP 8, SNP 12 or SNP 13. The number of transmitted or non-transmitted haplotypes from heterozygous parents to their affected offspring is shown. The statistic used to evaluate transmission distortion is PDT⁸. Haplotypes shown in bold are discussed in the text.

* Includes all haplotypes where none of the rare alleles of SNP8, 12 or 13 are present.

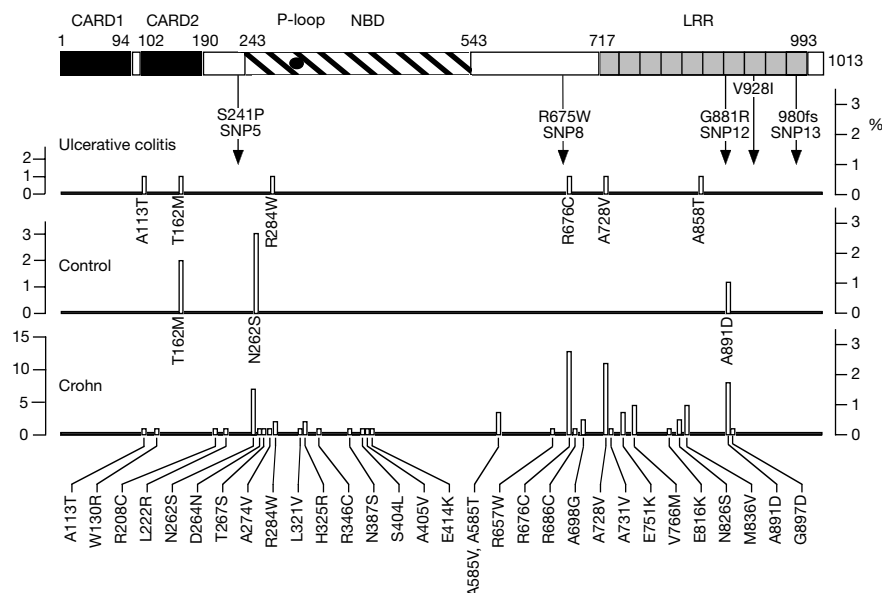


Figure 2 Representation of the IBD1/NOD2 protein variants. The translation product deduced from the cDNA sequence of the candidate *IBD1* gene is identical to that of *NOD2* (ref. 14). The polypeptide contains two caspase recruitment domains (CARD), a nucleotide-binding domain (NBD) and ten 27-amino-acid, leucine-rich repeats (LRRs). Black circle indicates the consensus sequence of the ATP/GTP-binding site motif A (P-loop) of the NBD. The sequence changes encoded by the three main variants associated with CD are SNP 8 (R675W), SNP 12 (G881R) and SNP 13 (980 frameshift). This frameshift changes a leucine to a proline at position 980, and is immediately followed by a

disease (IBD) in mice has been associated with mutations in Toll-like receptor 4 (TLR4)—a member of a family of NF- κ B activators that is known to bind LPS through its LRR domain^{19,20}. Second, antibiotic therapy causes transient improvement of CD patients, supporting the hypothesis that enteric bacteria may have an aetiological role in CD²¹. Third, NF- κ B has a pivotal role in IBD and is activated in mononuclear cells of the intestinal lamina propria in CD²². Last, CD treatment is based on the use of sulphasalazine and glucocorticoids—two known NF- κ B inhibitors^{23,24}.

Table 3 Allele and genotype frequencies of the three variants associated with CD

Frequencies of the three rare variant alleles					
	Number of chromosomes	SNP 8	SNP 12	SNP 13	Total
Unaffected	206	0.04	0.01	0.02	0.07
UC patients	318	0.03	0.00	0.01	0.05
CD patients	936	0.11	0.06	0.12	0.29
Distribution of variant genotypes and associated risks					
	Genotype				
	No variant	Simple heterozygous	Homozygous	Compound heterozygous	
Distribution					
Unaffected	88	15	0	0	
UC patients	145	13	1	0	
CD patients	267	133	28	40	
Risk for CD					
Relative risk	1	3	38	44	
Absolute risk	7×10^{-4}	2×10^{-3}	3×10^{-2}	3×10^{-2}	

Genotypes of patients: simple heterozygous, presence of a single rare variant; homozygous, presence of the same variant on both chromosomes 16; compound heterozygous, presence of two different variants; and no variant. Risk of CD for each genotype is computed assuming a prevalence of one per 1,000 and Hardy-Weinberg equilibrium for these markers in the general population. The three rare variants of SNP 8, 12 and 13 were never observed on the same haplotype. UC, ulcerative colitis.

stop codon. SNP 5 is described in Table 1. The allele frequencies of the V928I polymorphism were not significantly different (0.92:0.08) in the three groups, and the corresponding genotypes were in Hardy-Weinberg equilibrium. The positions of the rarer missense variants, observed in 457 CD patients, 159 ulcerative colitis patients and 103 unaffected unrelated individuals, are indicated for these groups. Left scale indicates the number of each identified variant in the investigated groups; right scale measures the mutation frequency.

Genetic susceptibility to CD is not limited to chromosome 16 and at least five additional loci have been implicated^{25–29}. The recognition of a transduction pathway that, when dysregulated, contributes to the pathogenesis of CD will accelerate the discovery of additional susceptibility genes. It will also contribute to the identification of associated environmental factors and focus the search for specific therapies. □

Methods

Families, microsatellite markers and contig construction

A total of 235 CD families (117 simplex nuclear families, 96 multiplex nuclear families, and 22 extended pedigrees, corresponding to a total of 179 CD patients and 261 unaffected relatives) was progressively recruited according to published diagnostic criteria³⁰. In addition, 100 multiplex and 59 simplex ulcerative colitis families were recruited from the same hospitals. Written informed consent was obtained from all participants. All relatives from 77 multiplex families were typed for 26 mapped microsatellite markers with an average resolution of 1 cM between SPN and D16S408. We constructed contigs using seven previously localized sequence tag sites (STSs; D16S541, D16S3035, D16S3136, D16S3117, D16S770, D16S416, D16S2623) and subsequently eight additional ones (wi-9288, wi-16305, shgc-17274, sgc-31023, sgc-32374, stSG-30035, wi-5812, D16S766) and 79 new STSs derived from the end sequences of the isolated BAC clones⁶.

Clones, sequencing and SNPs

The DNA of BAC clone hb87b10 containing D16S3136 was fragmented by sonication and subcloned in bacteriophage M13. We used sequences from both strands of 706 subclones and from direct primer walking to reconstruct the initial BAC sequence using PhredPhrap (<http://www.phrap.org>). Identity search in DNA databases identified two overlapping sequenced BACs (AC007334, GenBank; AC007728, GenBank). Homology search, performed on the extended sequence with BLAST v1.4 in GenBank release 114, identified 10 Unigene clusters. The following EST clones corresponding to some of these clusters were obtained from the American Type Culture Collection (<http://www.atcc.org>) and sequenced completely to identify additional transcribed regions: AI125217, AA417810, AI375427, AA021341, AI090427, AA910520, AA731089. Clones AI090427 and AA910520, corresponding to hs135201, were used to screen a blood leukocyte cDNA library (no. 938202; Stratagene), and retrieved 11 clones of the *IBD1* candidate gene.

A total of 123 STSs, mostly selected from putatively transcribed sequences (EST homologies and GRAIL v2.0 predicted exons), including 11 exons of KIAA0849, was

sequenced following amplification from the DNA of ten CD patients and two unaffected individuals. Of 35 identified SNPs, SNP 1–11, selected for their rare-allele frequencies greater than 0.08, were typed on 1,272 members of the 235 CD families. SNP 12 and 13 were further identified by sequencing the 11 exons of the candidate *IBD1* gene in 50 CD patients (SNP 1–13, GenBank accession numbers G67943–G67955, are submitted to dbSNP of the National Center for Biotechnology Information) and typed on the same group of individuals. To search for rare variant alleles, we subsequently investigated the 11 exons of 457 CD patients, 159 ulcerative colitis patients and 103 unaffected unrelated individuals. All variant alleles were confirmed by sequencing a second independent amplification product.

Data analysis

Genotypic data were analysed for linkage using the NPL score of GeneHunter v2.0. Data from linkage disequilibrium mapping of CD were analysed initially with the transmission disequilibrium test⁷ using a single trio (one affected and both parents) per family. Subsequently, the pedigree disequilibrium test was performed using the PDT 2.11 program⁸ to analyse data from all family relatives. We estimated allele frequencies for 3 groups, 418 unrelated CD patients, 159 ulcerative colitis patients and 103 controls (including 78 unaffected, unrelated spouses of CD patients and 25 unrelated CEPH family members).

Received 5 February; accepted 30 March 2001.

1. Calkins, B. M. & Mendelhoff, A. I. In *Inflammatory Bowel Disease* (eds Kirsner, J. B. & Shorter, R. G.) 31–68 (Williams & Wilkins, Baltimore, 1995).
2. Hugot, J. P., Zouali, H., Lesage, S. & Thomas, G. Etiology of the inflammatory bowel diseases. *Int. J. Colorectal Dis.* **14**, 2–9 (1999).
3. Hugot, J. P. *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* **379**, 821–823 (1996).
4. Olavesen, M. G. *et al.* Analysis of single-nucleotide polymorphisms in the interleukin-4 receptor gene for association with inflammatory bowel disease. *Immunogenetics* **51**, 1–7 (2000).
5. Hugot, J. P. *et al.* Mutation screening in the CD19 and CD43 (sialophorin) genes in Crohn Disease patients. *Gastroenterology* **116**, A740 (1999).
6. Zouali, H. *et al.* Refined mapping of the inflammatory bowel disease 1 gene. *Eur. J. Hum. Genet.* (submitted).
7. Spielman, R. S., McGinnis, R. E. & Ewens, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
8. Martin, E. R., Monks, S. A., Warren, L. L. & Kaplan, N. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* **67**, 146–154 (2000).
9. Kuster, W., Pascoe, L., Purmann, J., Funk, S. & Majewski, F. The genetics of Crohn's disease: complex segregation analysis of a family study with 265 patients with Crohn's disease and 5387 relatives. *Am. J. Med. Genet.* **32**, 105–108 (1989).
10. Monsen, U., Iselius, L. & Hellers, G. Evidence for a recessive gene in Crohn's disease. *Acta Chir. Scand.* **559** (suppl.), 7–42 (1990).
11. Orholm, M. *et al.* Investigation of inheritance of chronic inflammatory bowel disease by complex segregation analysis. *Br. Med. J.* **306**, 20–24 (1993).
12. Colombel, J. F. *et al.* Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* **111**, 604–607 (1996).
13. The IBD International Genetics Consortium. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *Am. J. Hum. Genet.* **68**, 1165–1171 (2001).
14. Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J. Biol. Chem.* **276**, 4812–4818 (2001).
15. Inohara, N., Ogura, Y., Chen, F. E., Muto, A. & Nuñez, G. Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *J. Biol. Chem.* **276**, 2551–2554 (2001).
16. Inohara, N. *et al.* Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor κ B. *J. Biol. Chem.* **274**, 14560–14567 (1999).
17. Bertin, J. *et al.* Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF- κ B. *J. Biol. Chem.* **274**, 12955–12958 (1999).
18. Inohara, N. *et al.* An induced proximity model for NF- κ B activation in the Nod1/RICK and RIP signaling pathways. *J. Biol. Chem.* **275**, 27823–27831 (2000).
19. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
20. Sundberg, J. P., Elson, C. O., Bedigian, H. & Birkenmeier, E. H. Spontaneous, heritable colitis in a new substrain of C3H/HeJ mice. *Gastroenterology* **107**, 1726–1735 (1994).
21. McKay, D. M. Intestinal inflammation and the gut microflora. *Can. J. Gastroenterol.* **13**, 509–516 (1999).
22. Schreiber, S., Nikolaus, S. & Hampe, J. Activation of nuclear factor κ B in inflammatory bowel disease. *Gut* **42**, 477–484 (1998).
23. Auphan, N., DiDonato, J. A., Rosette, C., Helmsberg, A. & Karin, M. Immunosuppression by glucocorticoids: inhibition of NF- κ B activity through induction of I κ B synthesis. *Science* **270**, 286–290 (1995).
24. Wahl, C., Liptay, S., Adler, G. & Schmid, R. M. Sulfasalazine: a potent and specific inhibitor of nuclear factor κ B. *J. Clin. Invest.* **101**, 1163–1174 (1998).
25. Satsangi, J. *et al.* Two stage genome wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosome 3, 7 and 12. *Nature Genet.* **14**, 199–202 (1996).
26. Hampe, J. *et al.* Linkage of inflammatory bowel disease to Human chromosome 6p. *Am. J. Hum. Genet.* **65**, 1647–1655 (1999).
27. Cho, J. H. *et al.* Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Hum. Mol. Genet.* **9**, 1425–1432 (2000).
28. Duerr, R. H., Barnada, M. M., Zhang, L., Pfutzer, R. & Weeks, D. E. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11–12. *Am. J. Hum. Genet.* **66**, 1857–1862 (2000).

29. Rioux, J. D. *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am. J. Hum. Genet.* **66**, 1863–1870 (2000).
30. Lennard-Jones, J. E. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol.* **170** (suppl.), 2–6 (1989).

Acknowledgements

We acknowledge patients and their families, and thank family recruitment doctors: J. Balanzo, B. Bonaz, Y. Bouhnik, G. Cadot, S. Cucchiara, B. Crusius, J. J. Delchier, B. Duclos, J. L. Dupas, J. P. Galmiche, J. P. Gendre, D. Golfain, C. Grännö, D. Heresbach, A. Lachaux, H. Lautraite, C. Lenaerts, E. Lerebours, V. Levy, R. Löfberg, H. Malchow, P. Marteau, A. Morali, F. Pallone, S. Pena, A. Rotenberg, I. Rousseau, J. Schmitz, F. Shanahan, I. Sobhani, H. Svensson, A. Van Gossom, M. Van Winkel and M. Veyrac. For assistance we are grateful to J. C. Beaudoin, F. Chareyre, C. Giudicelli, T. Hung Bui, M. Legrand, A. Marcadet, A. Martins, C. de Toma, E. Tubacher. We thank H. Cann for critically reading the manuscript. This project received support from European Community, MENRT, INSERM, Direction Générale de la Santé, Association François Aupetit, IRMAD and the Swedish Society of Medicine.

Correspondence and requests for materials should be addressed to G.T. (e-mail: thomas@cephb.fr).

A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease

Yasunori Ogura^{*†}, Denise K. Bonen^{‡†}, Naohiro Inohara^{*}, Dan L. Nicolae[§], Felicia F. Chen^{*}, Richard Ramos[‡], Heidi Britton[‡], Thomas Moran[‡], Reda Karaliuskas[‡], Richard H. Duerr^{||}, Jean-Paul Achkar[†], Steven R. Brant[#], Theodore M. Bayless[#], Barbara S. Kirschner[☆], Stephen B. Hanauer[‡], Gabriel Nuñez^{*††} & Judy H. Cho^{‡††}

^{*} Department of Pathology and Comprehensive Cancer Center, The University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

[‡] The Martin Boyer Laboratories, Gastroenterology Section, Department of Medicine, The University of Chicago Hospitals, Chicago, Illinois 60637, USA

[§] Department of Statistics, and [☆] Department of Pediatrics, University of Chicago, Chicago, Illinois 60637, USA

^{||} Department of Medicine and Center for Genomic Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

[¶] Department of Gastroenterology, The Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

[#] The Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[†] These authors contributed equally to this work

^{††} These authors share senior authorship

Crohn's disease is a chronic inflammatory disorder of the gastrointestinal tract, which is thought to result from the effect of environmental factors in a genetically predisposed host. A gene location in the pericentromeric region of chromosome 16, *IBD1*, that contributes to susceptibility to Crohn's disease has been established through multiple linkage studies^{1–6}, but the specific gene(s) has not been identified. *NOD2*, a gene that encodes a protein with homology to plant disease resistance gene products is located in the peak region of linkage on chromosome 16 (ref. 7). Here we show, by using the transmission disequilibrium test and case-control analysis, that a frameshift mutation caused by a cytosine insertion, 3020insC, which is expected to encode a truncated *NOD2* protein, is associated with Crohn's disease. Wild-type *NOD2* activates nuclear factor NF- κ B, making it responsive to bacterial lipopolysaccharides; however, this induction was deficient in mutant *NOD2*. These results implicate *NOD2* in susceptibility to Crohn's disease, and suggest a link between an

innate immune response to bacterial components and development of disease.

The idiopathic inflammatory bowel diseases (IBDs), which include Crohn's disease (CD) and ulcerative colitis, are chronic disorders of the gastrointestinal tract with unknown aetiology, and with a combined prevalence of about 150–200 cases per 100,000 in western countries^{8,9}. Although the aetiology of IBD is unknown, an abnormal inflammatory response directed against enteric microflora in a genetically susceptible host has been proposed¹⁰. Familial clustering of disease and studies of twins strongly suggest that IBD, and in particular CD, is a genetic disorder¹¹. Genome-wide searches for IBD-susceptibility genes have resulted in the identification of several loci for CD and/or ulcerative colitis, most notably for CD, in the pericentromeric region of chromosome 16 (*IBD1*)^{1–6}.

NOD2 has structural homology with both the apoptosis regulators Apaf-1/Ced-4 and a class of plant disease resistant (R) gene products⁷. Like the latter gene products, *NOD2* comprises an amino-terminal effector domain, a nucleotide-binding domain and leucine-rich repeats (LRRs) (ref. 7). *NOD2* has been mapped to chromosome 16q12 (ref. 7) and is tightly linked to markers D16S3396, D16S416 and D16S419 (Fig. 1a)—a site that precisely overlaps with *IBD1* (ref. 1). Given the genomic localization and the role of NOD proteins in recognizing bacterial components¹², we thought that *NOD2* might function as a susceptibility gene for CD.

The 12-exon genomic organization of the *NOD2* gene was determined by aligning the complementary DNA sequence

(AF178930) with one genomic bacterial artificial chromosome (BAC) clone, RP11-327F22 (AC007728) (Fig. 1a). All coding exons and flanking introns were sequenced in 12 affected individuals from pure CD families with increased linkage scores at D16S3396, as well as in 4 case controls. In three CD patients, a cytosine insertion was observed in exon 11 at nucleotide 3020 (3020insC) (Fig. 1b). 3020insC resulted in a frameshift at the second nucleotide of codon 1007 (Fig. 1b), and a Leu1007→Pro substitution in the tenth LRR, followed by a premature stop codon (Fig. 1c). The predicted truncated *NOD2* protein contained 1,007 amino acids instead of the 1,040 amino acids of the wild-type *NOD2* protein (Fig. 1d).

We used an allele-specific polymerase chain reaction (PCR) assay (Fig. 2) to type 3020insC in IBD families and case controls. Analysing only one CD patient per independent family, we observed preferential transmission (Table 1) from heterozygous parents to affected children of 3020insC (39 transmissions and 17 non-transmissions; $P = 0.0046$). Analysing all independent nuclear families by sib-TDT (ftp://lahmed.stanford.edu/pub/aspx/index.html), the empirical P value was similar, $P = 0.0007$. As expected from linkage studies^{2–5}, no preferential transmission of 3020insC was observed among families with ulcerative colitis (data not shown). As 365 of the 416 independent CD families have several affected individuals, the applicability of these associations to the more common, sporadic cases requires further study.

Additional support for association to CD was provided by case-

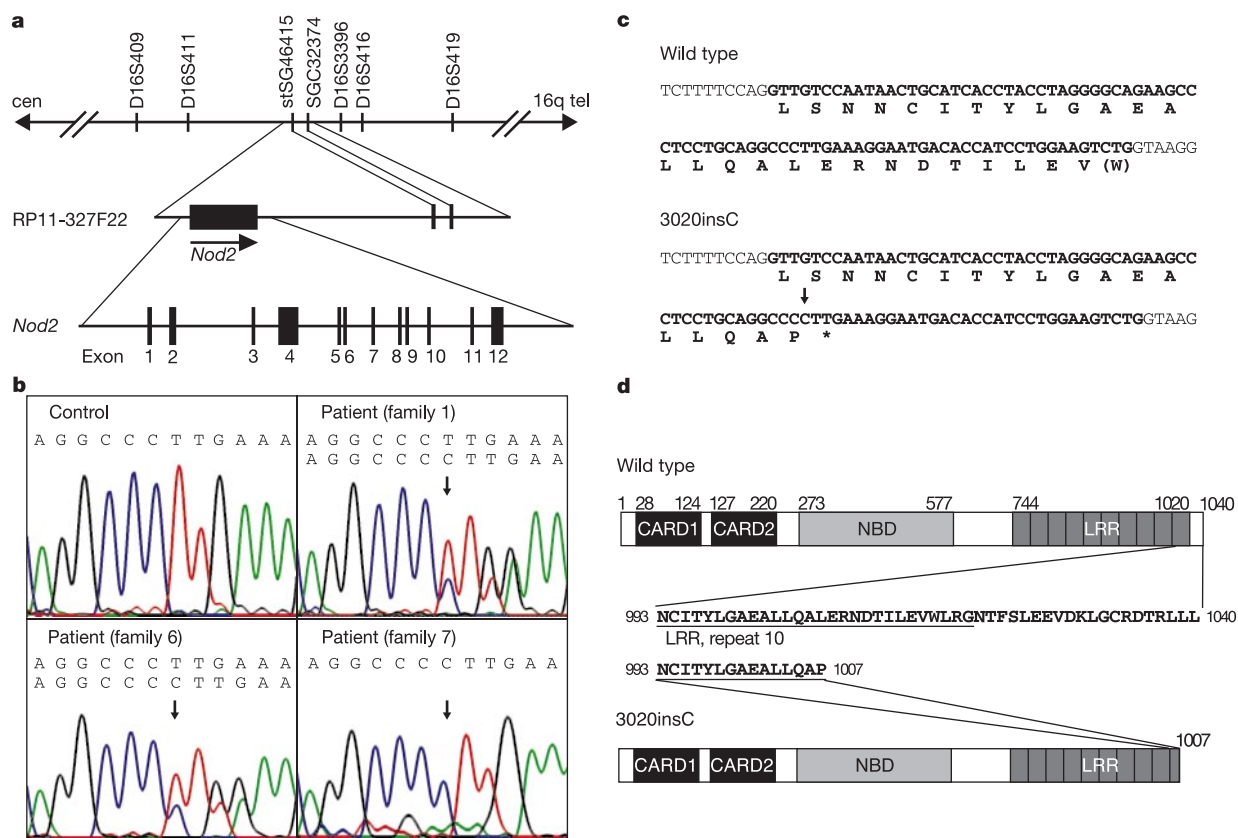


Figure 1 Identification of a frameshift *NOD2* mutation in affected individuals from CD families. **a**, Physical map of the region of interest at 16q12. Approximate positions of chromosomal and genetic markers are based on ref. 22. The human genomic BAC clone RP11-327F22 contains the *NOD2* gene and markers sSG46415 and SGC32374. The exon–intron organization of the human *NOD2* gene is shown underneath. **b**, DNA sequence electropherograms (exon 11) from control and three affected individuals from CD families. Patients from families 1 and 6 are heterozygous, whereas the patient from family 7 is homozygous for 3020insC. The cytosine insertion is indicated by an arrow.

c, Nucleotide and predicted amino-acid sequence of exon 11 and flanking introns from wild-type control and patients with 3020insC. The exon sequence is shown in bold. The site of 3020insC is indicated by an arrow. Residue (W) indicates that a nucleotide from exon 12 contributes to the codon. **d**, Domain structure of *NOD2*, illustrating the site of protein truncation. Caspase-recruitment domains (CARDs), the nucleotide-binding domain (NBD) and ten LRRs are shown. Residues of the tenth LRR are underlined. Numbers indicate residue positions.

Table 1 TDT demonstrates preferential transmission of the 3020insC to CD patients

Source	One CD patient per family			All CD patients	
	Transmitted	Not transmitted	P value	Transmitted	Not transmitted
Univ. of Chicago	21	10	0.0046	32	16
Johns Hopkins	4	4		10	8
Univ. of Pittsburgh	14	3		26	9
Total	39	17		68	33

control analysis, in which, using one CD individual per independent family, the 3020insC allele frequency among all CD groups was 8.2% (Table 2). The allele frequencies of 3020insC were comparable among Jewish (8.4%) and non-Jewish Caucasians (8.1%). Among case controls (Table 2), the allele frequency in four separate Caucasian cohorts of 4.0% was significantly lower than in CD patients ($P = 0.0018$, by large-sample approximations to a two-sample binomial test). The allele frequency of the 3020insC among 182 unrelated ulcerative colitis patients was 3.0%, and was significantly lower than the frequency among CD patients ($P = 0.0010$). The genotype frequencies of 3020insC in unrelated CD individuals were 11 homozygotes, 46 heterozygotes and 359 wild-type homozygotes.

Among case controls, there were 23 heterozygous individuals, with the remaining being wild-type homozygotes. The genotype-relative risk (GRR) for heterozygous and homozygous 3020insC was 1.5 and 17.6, respectively, as compared with wild-type controls. Given its frequency, 3020insC is unlikely to account completely for the observed evidence of linkage at *IBD1*, and other variants of *NOD2* may confer additional disease risk. For example, two single-nucleotide polymorphisms in *NOD2* have been identified, 2722G→C (Gly908Arg) and 2104C→T (Arg702Trp), which are highly associated with CD by the transmission disequilibrium test (data not shown). Furthermore, other susceptibility genes might also be present in this broad region^{1–6} of linkage on chromosome 16.

NOD2 has been shown to activate NF-κB and to confer responsiveness to bacterial lipopolysaccharides^{7,12}. To test the ability of wild-type and mutant *NOD2* to activate NF-κB, human embryonic kidney (HEK) 293T cells were transiently co-transfected with wild-type or 3020insC plasmids and an NF-κB reporter construct. In the absence of lipopolysaccharide (LPS), expression of both wild-type and mutant *NOD2* induced activation of NF-κB (Fig. 3a). Notably, equivalent levels of wild-type and mutant *NOD2* protein expression (as assessed by immunoblotting of total lysates) resulted in similar levels of NF-κB activation (Fig. 3a).

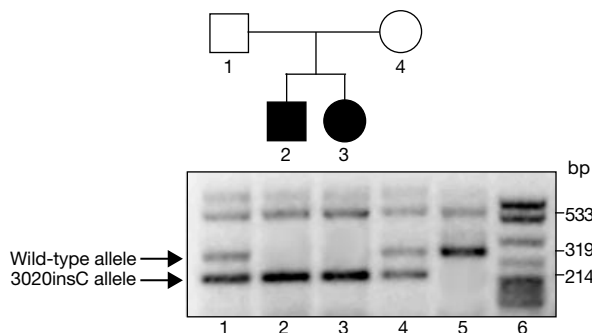


Figure 2 Determination of transmission of the 3020insC mutation in a CD family by allele-specific PCR. Multiplex PCR was used to generate a nonspecific 533-bp product, along with allele-specific amplicons: a 319-bp fragment (wild type) and a 214-bp fragment (3020insC). In this family, both parents (lanes 1 and 4) are heterozygous for 3020insC, whereas both children (lanes 2 and 3) have CD and are homozygous for 3020insC. Lane 5, wild-type control; lane 6, pBR322 DNA MspI markers. Numbers on the right indicate the size of fragments.

Table 2 Allele frequency of 3020insC in unrelated Crohn's disease patients and controls

Crohn's disease			Case controls		
Source	Sample size	3020insC*	Source	Sample size	3020insC*
Univ. of Chicago	212	7.3	Chicago	65	3.8
Johns Hopkins	88	6.8	Baltimore	46	3.2
Univ. of Pittsburgh	116	10.8	San Francisco	81	3.1
			Germany	94	5.3
Total	416	8.2		287	4.0

For analysis, one CD patient was selected from 416 independent families; the difference in allelic frequency between CD patient and control was significant ($P = 0.0018$, by the large sample approximations to a two sample binomial test).

* Per cent allele frequency.

Like *NOD2*, cytosolic plant disease resistant proteins have carboxy-terminal LRRs that are critical for the recognition of pathogen components and induction of pathogen-specific responses^{13–15}. We therefore compared the ability of wild-type and mutant *NOD2* proteins to induce NF-κB activity in response to LPS. Because overexpression of *NOD2* induces potent NF-κB activation (Fig. 3a), we transfected the cells with low amounts of wild-type and mutant

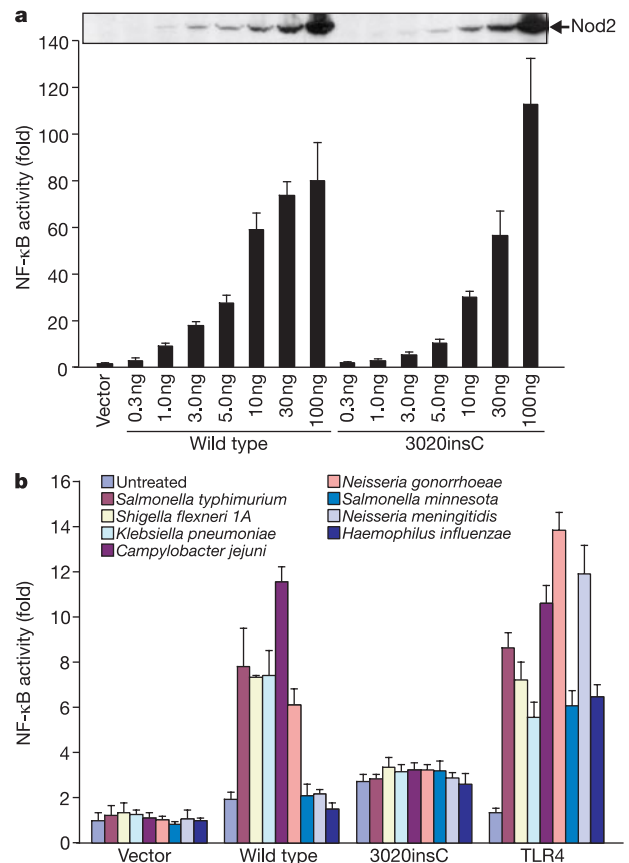


Figure 3 Differential responsiveness of wild-type and mutant *NOD2* to LPS. **a**, HEK293T cells were co-transfected in triplicate with the indicated amounts of pcDNA3 (vector), wild-type pcDNA3-*NOD2*, or pcDNA3-*NOD2* 3020insC and pEF-BOS-β-gal and pBVI-luc reporter plasmids. Values represent means ± s.d. Expression of wild-type and mutant *NOD2* proteins in cell extracts is shown on top. **b**, HEK293T cells were co-transfected in triplicate with 0.3 ng of pcDNA3-*NOD2*, 3 ng of pcDNA3-*NOD2* 3020insC, 3 ng of pcDNA3-TLR4 plus 3 ng of pcDNA3-MD-2 (indicated by TLR4) or pcDNA3 (vector) and pEF-BOS-β-gal and pBXIV-luc. Under these conditions, both wild-type and mutant *NOD2* constructs induced similar levels of basal NF-κB activity. Eight hours after transfection, cells were treated with 10 μg ml⁻¹ of LPS from indicated bacteria. Values represent means ± s.d. Results are representative of at least five independent experiments.

NOD2 plasmids to induce similar levels of protein expression and basal NF- κ B activity (Fig. 3a). LPS from various bacteria induced NF- κ B activation in cells expressing wild-type NOD2, but not in cells transfected with control plasmid (Fig. 3b).

Significantly, the ability of mutant NOD2 to confer responsiveness to LPS was greatly diminished when compared with wild-type NOD2 (Fig. 3b). Differential regulation of NOD2 function by LPS from different bacteria was observed (Fig. 3b), whereas all LPS preparations induced NF- κ B activation comparably in cells transfected with Toll-like receptor-4 (TLR-4), a cell-surface receptor for LPS¹⁶.

The innate immune system regulates the immediate response to microbial pathogens and is initiated by recognition of specific pathogen components by receptors in host immune cells¹⁶. NOD1 and NOD2 seem to function as intracellular receptors for LPS with the LRRs required for responsiveness¹². We have shown here that truncation of the tenth LRR of NOD2 is associated with CD. Consistent with earlier linkage studies^{2,5}, this variant is associated solely with CD, and not with ulcerative colitis. Functional analyses indicate that the disease-associated NOD2 variant is significantly less active than the wild-type protein in conferring responsiveness to bacterial LPS. In plant NOD2 homologues, the LRRs determine the specificity for pathogen products and alterations in LRRs can result in unresponsiveness to particular pathogens^{13–15}. Similarly, genetic variation in the LRRs of TLR4 account for inter-individual differences in bronchial responsiveness to aerosolized LPS¹⁷.

Several mechanisms can be envisaged to account for susceptibility to CD in individuals carrying this variant. NOD2 is a cytosolic protein whose expression is restricted to monocytes, with no expression detected in lymphocytes⁷. A deficit in sensing bacteria in monocytes/macrophages might result in an exaggerated inflammatory response by the adaptive immune system. A related possibility is that wild-type NOD2 may mediate the induction of cytokines such as interleukin-10 that can downregulate the inflammatory response^{18,19}. Finally, variation in the LRRs of plant NOD2 homologues can result in recognition of new specificities for pathogen components^{13,14}. Thus, it is also possible that NOD2 variants might act as gain-of-function alleles for unknown pathogens. Our studies implicate NOD2 in susceptibility to CD, and suggest a link between an innate response to bacterial components and development of disease. □

Methods

IBD families

IBD families were ascertained for linkage and association studies (affected child with both parents) through the University of Chicago, the Johns Hopkins Hospital and the University of Pittsburgh. In all cases informed consent for a molecular genetic study was obtained, and the study protocol was approved by the individual institutional review boards.

Allele-specific PCR

We used primers framing a 533-base-pair region surrounding the 3020insC allele to amplify genomic DNA isolated from controls and patients by PCR (sense, 5'-CTGAGCCTTTGTGATGAGC-3'; antisense, 5'-TCTTCAACCACATCCCCATT-3'). In addition, each PCR reaction contained two additional primers designed to detect the wild-type allele (sense, 5'-CAGAAGCCCTCCTGCAGGCCCT-3') and another primer designed to detect the 3020insC allele (antisense, 5'-CGCGTGTTCATTCCTTTCATGGGGC-3'). The 3020insC was confirmed by DNA sequencing. We performed multiplex PCR with all four primers in one tube. PCR products were isolated on 2% agarose gels and visualized with ethidium bromide.

Data analysis

The *P* values for the TDT test²⁰ were calculated using a binomial exact test. Simulations (500,000 replicates) were done using the sib-TDT software (<http://lahmed.stanford.edu/pub/aspex/index.html>) to calculate empirical probabilities for the TDT χ^2 statistic when all independent nuclear families were counted. This calculation was done by permuting parent alleles while fixing the IBD status of siblings within a family. We estimated the frequency of the genotypes in the affected individuals from 416 unrelated CD patients. The GRRs²¹ are defined as the ratio of the marginal penetrance of the 3020insC homozygote and heterozygote genotypes to the wild-type homozygotes. Using Bayes rule, the GRRs can be expressed as a function of the allele frequencies in the case and control groups. For the

control group, we assumed that the alleles are in Hardy–Weinberg equilibrium.

Expression plasmids and immunoblotting

The expression plasmids pcDNA3-NOD2, pcDNA3-TLR4 and pcDNA3-MD-2 have been described^{7,12}. The expression plasmid producing the NOD2 Δ 33 mutant (3020insC) was generated by PCR and cloned into pcDNA3 (Invitrogen), and confirmed by DNA sequencing. Expression of untagged NOD2 proteins in transfected cells was determined by immunoblotting using affinity-purified rabbit anti-NOD2 antibody, as described⁷. To raise the antibody, we overexpressed recombinant NOD2 protein (residues 28–301) in *Escherichia coli* strain BL21(DE3) using the pET-30a vector (Novagen). Recombinant NOD2 protein containing a C-terminal histidine tag was purified using a nickel column (Novagen) and injected into rabbits.

NF- κ B activation assay

We carried out NF- κ B activation assays as described^{7,12}. Briefly, HEK293T cells were co-transfected with 12 ng of the reporter construct pBVI-Luc, the indicated amounts of each expression plasmid and 120 ng of pEF-BOS- β -gal in triplicate in the presence or absence of LPS^{7,12}. LPS from various sources were obtained from Sigma or from several investigators. Twenty-four hours after transfection, cell extracts were prepared and the relative luciferase activity was measured as described^{7,12}. Results were normalized for transfection efficiency with values obtained with pEF-BOS- β -gal.

Received 18 January; accepted 30 March 2001.

- Hugot, J. P. *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* **379**, 821–823 (1996).
- Ohmen, J. D. *et al.* Susceptibility locus for inflammatory bowel disease on chromosome 16 has a role in Crohn's disease, but not in ulcerative colitis. *Hum. Mol. Genet.* **5**, 1679–1683 (1996).
- Curran, M. E. *et al.* Genetic analysis of inflammatory bowel disease in a large European cohort supports linkage to chromosomes 12 and 16. *Gastroenterology* **115**, 1066–1071 (1998).
- Cavanaugh, J. A. *et al.* Analysis of Australian Crohn's disease pedigrees refines the localization for susceptibility to inflammatory bowel disease on chromosome 16. *Ann. Hum. Genet.* **62**, 291–298 (1998).
- Cho, J. H. *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc. Natl Acad. Sci. USA* **95**, 7502–7507 (1998).
- Annese, V. *et al.* Genetic analysis in Italian families with inflammatory bowel disease supports linkage to the IBD1 locus—a GISC study. *Eur. J. Hum. Genet.* **7**, 567–573 (1999).
- Ogura, Y., Inohara, N., Benito, A., Chen, F. F. & Núñez, G. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J. Biol. Chem.* **276**, 4812–4818 (2001).
- Loftus, E. V. Jr *et al.* Crohn's disease in Olmsted County, Minnesota, 1940–1993: incidence, prevalence, and survival. *Gastroenterology* **114**, 1161–1168 (1998).
- Kyle, J. Crohn's disease in the northeastern and northern Isles of Scotland: an epidemiological review. *Gastroenterology* **103**, 392–399 (1992).
- Fiocchi, C. Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* **115**, 182–205 (1998).
- Binder, V. Genetic epidemiology of inflammatory bowel disease. *Digest. Dis.* **16**, 351–355 (1998).
- Inohara, N., Ogura, Y., Chen, F. F., Muto, A. & Núñez, G. Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *J. Biol. Chem.* **276**, 2551–2554 (2001).
- Parniske, M. *et al.* Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. *Cell* **91**, 821–832 (1997).
- Ellis, J. G., Lawrence, G. J., Luck, J. E. & Dodds, P. N. Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell.* **11**, 495–506 (1999).
- Dixon, M. S., Golstein, C., Thomas, C. M., van Der Biezen, E. A. & Jones, J. D. Genetic complexity of pathogen perception by plants: the example of Rcr3, a tomato gene required specifically by Cf-2. *Proc. Natl Acad. Sci. USA* **97**, 8807–8814 (2000).
- Aderem, A. & Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782–787 (2000).
- Arbour, N. C. *et al.* TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nature Genet.* **25**, 187–191 (2000).
- Moore, K. W. *et al.* Interleukin-10. *Annu. Rev. Immunol.* **11**, 165–190 (1993).
- Berg, D. J. *et al.* Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4⁺ TH1-like responses. *J. Clin. Invest.* **98**, 1010–1020 (1996).
- Spielman, R. S. & Ewens, W. J. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**, 983–989 (1996).
- Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
- Han, C. S. *et al.* Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. *Genome Res.* **10**, 714–721 (2000).

Acknowledgements

We thank the patients and their families that participated in this study. We acknowledge the contributions of J. B. Kirsner and M. Boyer. We thank T. Ross, P. Lucas and L. McAllister-Lucas for critically reading the manuscript; and A. Moran and M. Apicella for LPS samples. The work was supported by grants from the NIH (G.N. and J.H.C.), the Crohn's and Colitis Foundation of America (J.H.C.), the Reva and David Logan Foundation (J.H.C.) and the Gastrointestinal Research Foundation (J.H.C.). Y.O. was supported by funds from Tokushima University, Japan. G.N. and J.H.C. contributed equally to this work and share senior authorship.

Correspondence and request for materials should be addressed to G.N. (e-mail: bclx@umich.edu).

The zebrafish Nodal signal Squint functions as a morphogen

Yu Chen & Alexander F. Schier

Developmental Genetics Program, Skirball Institute of Biomolecular Medicine, and Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA

Secreted morphogens induce distinct cellular responses in a concentration-dependent manner and act directly at a distance^{1–7}. The existence of morphogens during mesoderm induction and patterning in vertebrates has been highly controversial, and it remains unknown whether endogenous mesoderm inducers act directly as morphogens^{8–10}, function locally⁹ or act through relay mechanisms^{11–12}. Here we test the morphogen properties of Cyclops and Squint—two Nodal-related transforming growth factor- β signals required for mesoderm formation and patterning in zebrafish^{13–16}. Whereas different levels of both Squint and Cyclops can induce different downstream genes^{14,17–19}, we find that only Squint can function directly at a distance. These results indicate that Squint acts as a secreted morphogen that does not require a relay mechanism.

The level of Nodal signalling is a key factor in determining the cell fate induced in responding cells^{14,17–21}. We first tested whether Cyclops (Cyc) and Squint (Sqt) can act on responding cells at a distance, and whether the activation of different downstream genes is dependent on concentration and distance. To provide a local source of Sqt or Cyc, *sqt* or *cyc* RNA was injected, together with the lineage tracer fluorescein-dextran, into a single cell of 128–256-cell zebrafish (*Danio rerio*) embryos (Fig. 1a). The injected embryos were fixed 3 h later before the onset of gastrulation (50% epiboly), and analysed by *in situ* RNA hybridization with antisense probes for *no tail/Brachyury* (*ntl*), *goosecoid* (*gsc*), *bhikhari* (*bik*) and *floating head* (*flh*)—known targets of Nodal signalling^{13,18,22–23}.

We found that *flh*, *ntl* and *bik* were expressed by cells 4–6 (*flh*) and 6–8 (*ntl*, *bik*) cell diameters away from Sqt-producing cells (Fig. 1c, e, m). In contrast, *gsc* was detected only in cells expressing *sqt* and their immediate neighbours (Fig. 1d). This pattern recapitulates the expression of these genes in the zebrafish organizer, where high levels of Nodal signalling are required to activate *gsc*, and lower levels are sufficient to induce expression of *flh*, *ntl* and *bik*¹⁸. When we reduced the levels of *sqt* tenfold, *ntl* was induced only close to the source and *gsc* was not induced (Fig. 1f, g). Expression of different levels of *cyc* resulted in a similar concentration-dependent induction of *ntl* and *gsc*. High levels induce the expression of both genes, whereas lower levels induce only *ntl* (Fig. 1i, j; and data not shown); unexpectedly, however, we never observed *ntl* or *bik* expression beyond two cell diameters from *cyc*-expressing cells, although the level of *cyc* was high enough to induce *gsc* expression (Fig. 1i, j, l).

Notably, increasing the injected *cyc* RNA to 60 pg did not change the ectopic *ntl* domain, but induced ectopic *gsc* expression in all cells injected with RNA (data not shown). This indicates that more active Cyc protein is made, although western analysis with a haemagglutinin A (HA)-tagged form of Cyc revealed no significant difference in HA-Cyc protein production on injection of 6 pg or 60 pg (data not shown). This result indicates that post-transcriptional mechanisms such as translational efficiency or protein stability might contribute to the range of action of *cyc*.

The expression of *bik* in *cyc*, *sqt* and *cyc;sqt* double mutants is consistent with different ranges for endogenous Cyc and Sqt (Fig. 1o–s). In *cyc* mutants, Sqt is sufficient to activate *bik* 6–8 cells from the margin (Fig. 1q), whereas in *sqt* mutants, Cyc induces *bik* only locally (Fig. 1p). The amino-terminal prodomain of transforming growth factor (TGF)- β signals is thought to interact

with extracellular matrix and so may be responsible for the different range of action of Cyc and Sqt^{9,24}. We found, however, that the prodomains do not change the range of chimaeric Cyc-Sqt and Sqt-Cyc proteins (Fig. 1h, k), indicating that differences in the

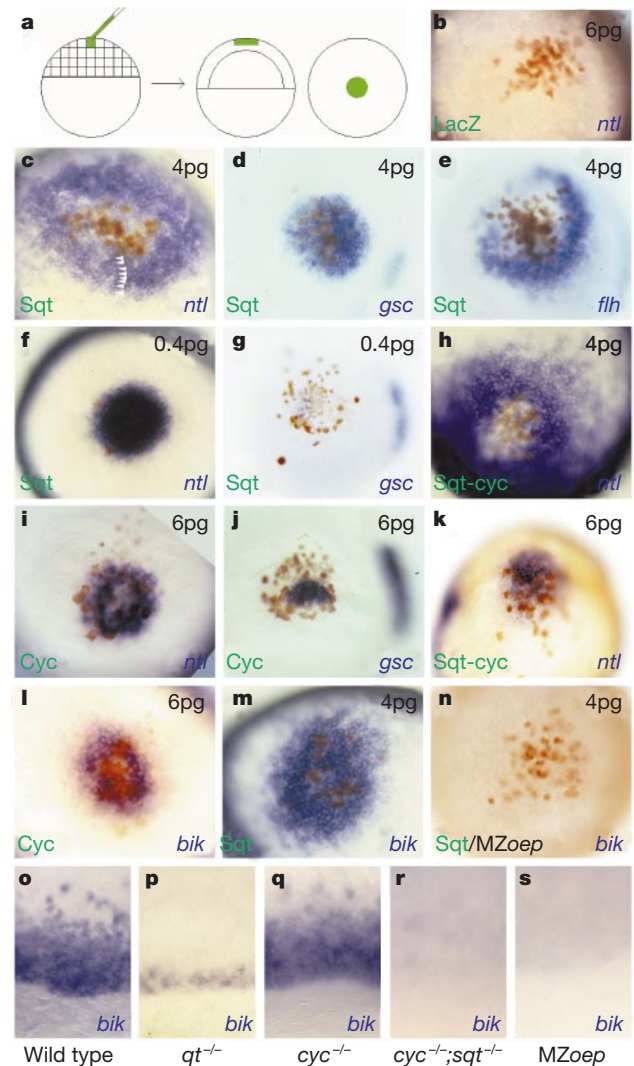


Figure 1 Squint has long-range effects, whereas Cyclops has short-range effects.

a, Single-cell injection; RNAs are co-injected with fluorescein-dextran (green) into wild-type embryos at the 128–256-cell stage; embryos are fixed 3 h later at 50% epiboly (left, lateral view; right, animal pole view). *ntl*, *gsc*, *flh* and *bik* expression is examined by whole-mount RNA *in situ* hybridization (blue stain); injected cells are stained brown by an anti-fluorescein antibody. **b–n**, Single-cell injections of RNA; all images are animal pole views. **b**, 6 pg of *lacZ* mRNA; no ectopic *ntl* expression is observed ($n = 12/12$). **c–e**, 4 pg of *sqt* RNA; *ntl* (**c**, $n = 27/27$) and *flh* (**e**, $n = 18/18$) are expressed in a wider domain than *gsc* (**d**, $n = 34/35$); arrowheads in **c** indicate the domain of ectopic *ntl* expression, about eight cell diameters from *sqt*-expressing cells (about 100 μ m). **f**, **g**, 0.4 pg of *sqt* RNA; the ectopic *ntl* domain almost overlaps with the Sqt-producing cells (**f**, $n = 14/15$); no ectopic *gsc* expression is induced (**g**, $n = 13/14$). **h**, 4 pg of *cyc-sqt* RNA ($n = 22/22$); the ectopic *ntl* domain is similar to that induced by *sqt* RNA (**c**). **i**, **j** and **l**, 6 pg of *cyc* RNA; the ectopic *ntl* and *bik* domains almost overlap with the Cyc-producing cells (**i**, $n = 18/20$; **l**, $n = 11/11$) even though the level of Cyc is high enough to induce ectopic *gsc* (**j**, $n = 16/18$). **k**, 6 pg of *sqt-cyc* RNA; the ectopic *ntl* domain is similar to that induced by *cyc* RNA injection ($n = 21/22$). **m**, **n**, 4 pg of *sqt* RNA in wild-type (**m**) and MZoeep (**n**) embryos; ectopic *bik* expression is induced up to eight cells away from the Sqt-producing cells in wild type (**m**, $n = 20/20$), but no ectopic *bik* expression is induced in MZoeep embryo (**n**, $n = 14/14$). **o–s**, *bik* expression at the blastoderm margin at 50% epiboly in different genetic backgrounds, lateral views; *bik* expression is dependent on Nodal signalling around the entire margin.

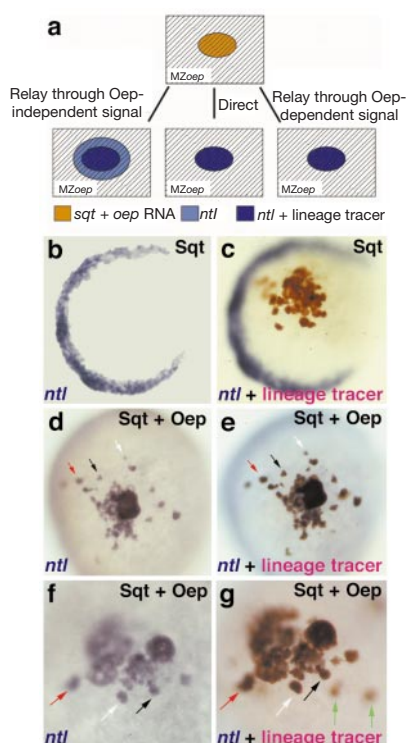


Figure 2 Sqt signals to distant cells either directly or through an Oep-dependent relay signal. **a**, Design and predictions for the experiment described in **b–g**. **b–g**, Single-cell injection into MZoop embryo of 4 pg of *sqt* RNA (**b, c**) or 4 pg of *sqt* RNA plus 4 pg of *oep* RNA (**d–g**). Note that MZoop mutants express *ntl* at the lateral and ventral margin, but not dorsally (**b, c**); **b, d, f** are *in situ* stainings of *ntl* expression (blue); **c, e, g** are double stainings of both *ntl* expression (blue) and lineage tracer detected with anti-fluorescein antibody (brown). **b, c**, No ectopic *ntl* expression (blue) is detected owing to the inability of MZoop cells to respond to Sqt ($n = 14/14$). **d–g**, *ntl* is only expressed by cells injected with *sqt* and *oep* RNA ($n = 19/19$). Arrows of different colours indicate the corresponding cells in **d** and **e**, as well as in **f** and **g**; note that there are two cells marked by the green arrows in **g** that contain the lineage tracer but do not express *ntl*.

mature region contribute to the different ranges of action. Together, these results indicate different effective ranges for Sqt and Cyc.

Given the dynamic cell movements during early embryogenesis, the apparent long-range effect of Sqt might be a result of the displacement of cells that initiated expression of *ntl* because early on they were adjacent to *sqt*-expressing cells²⁵. Although the short-range effect of Cyc argues against this possibility, we tested it by transplanting labelled wild-type cells at a distance from the *sqt*-expressing cells. To make sure that these two cell populations never got within close proximity of each other, we monitored their relative positions after transplantation. We found that transplanted cells that maintained their distance from the Sqt source expressed *ntl*, as did their neighbouring host cells (Supplementary Information Fig. 1). Thus, cell movement does not cause the long-range effect of Sqt.

Two models can explain the long-range effect of Sqt: Sqt might act as a morphogen and directly activate target gene expression at a distance; or Sqt might activate a relay signal, which then acts on distant cells. Both mechanisms have been invoked previously to account for the long-range effects of non-endogenous TGF- β signals^{8–12}. To assay the range of Sqt's direct action, we made use of the observation that cells that lack the activity of the EGF-CFC protein One-eyed pinhead (Oep) cannot perceive Nodal signals²². Oep is an extracellular cofactor required cell-autonomously for the reception of Nodal signals^{22,26–28}.

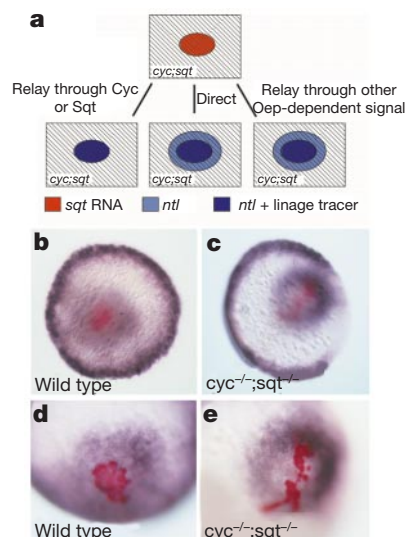


Figure 3 The long-range effect of Sqt does not depend on the induction of the endogenous *cyc* or *sqt* genes. **a**, Design and predictions for the experiments shown in **b–e**. **b–e**, Single-cell injection of 4 pg of *sqt* RNA plus lineage tracers biotin-dextran and rhodamine-dextran into wild-type (**b, d**, $n = 18/18$) and *cyc;sqt* double mutant embryos (**c, e**, $n = 12/12$) at the 128–256-cell stage. **b, c**, Focus is on the expression domain of *ntl* in marginal cells; *cyc;sqt* double mutants were distinguished from sibling embryos (**b**) by the lack of *ntl* expression on the dorsal side (**c**). **d, e**, High-magnification views of the embryos in **b** and **c**; the lineage tracer biotin-dextran was detected with the ABC kit from Vectastain (red).

Maternal-zygotic mutants for *oep* (MZoop embryos) have an identical phenotype to *cyc;sqt* double mutants and are unresponsive to Nodal signals²². These mutants are not impaired in signalling for fibroblast growth factor (FGF), bone morphogenetic protein (BMP) or Activin²², but none of the known Nodal-responsive genes (*bik*, *ntl*, *gsc*, *flh*, *antivin*, *pitx2*, *cyc*) can be activated by overexpression of Sqt in MZoop mutants, and ubiquitous or local expression of Sqt has no phenotypic effects in MZoop mutants²² (Figs 1n, 2c; and data not shown). Using these MZoop mutants, we tested whether Sqt acts directly at a distance or indirectly through a relay signal.

We first juxtaposed Sqt-responsive cells with Sqt non-responsive cells (*oep* mutant) to test whether an Oep-independent relay signal such as FGF¹² is involved in the long-range effect of Sqt (Fig. 2). We injected both *sqt* and *oep* RNAs into a single cell of 128–256-cell MZoop embryos. *oep* RNA injection allows the injected cells to transduce the Sqt signal. If Sqt can induce a relay signal whose transduction is Oep independent and activates target genes such as *ntl*, then *ntl* would be expressed not only in the injected cells but also in the surrounding *oep* mutant cells (Fig. 2a). But if Sqt acts as a morphogen or relays through signals whose transduction is dependent on Oep, then *ntl* expression would be observed only in cells expressing Oep. Consistent with the second scheme, we found that only cells containing both Oep and Sqt expressed *ntl* (Fig. 2d–g). No non-autonomous induction of *ntl* was observed.

To test whether the long-range effect of Sqt is a result of propagated induction of *cyc* or *sqt*, we analysed *ntl* expression induced by Sqt in *cyc;sqt* double-mutant embryos (Fig. 3a). We found that the domain of *ntl* expression induced in these embryos (Fig. 3c, e) is very similar to that in wild-type embryos (Fig. 3b, d), suggesting that the long-range effect of Sqt does not depend on the induction of the endogenous *cyc* or *sqt* genes.

These results are consistent with the idea that Sqt acts as a morphogen, but it is also possible that Sqt may relay through an unknown signal whose transmission is Oep dependent (Fig. 4a). To address this possibility, we injected *sqt* RNA into a single cell of

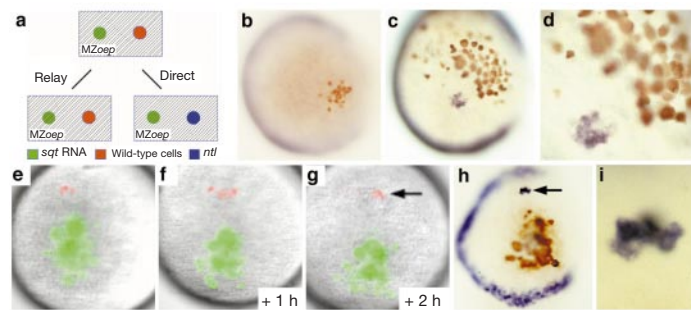


Figure 4 Sqt signals directly to distant cells. **a**, Experimental design and predictions. **b**, Control; transplantation of wild-type cells (brown) into MZ*oep* embryo, no ectopic *ntl* expression is observed ($n = 14/14$). Note that MZ*oep* mutants express *ntl* at the lateral and ventral margin, but not dorsally. **c–i**, Single-cell injection of 4 pg of *sqt* RNA and fluorescein-dextran (green) into MZ*oep* embryo, followed by transplanting wild-type donor cells labelled with biotin-dextran and rhodamine-dextran (red) at a distance to the *sqt*-

injected cells (green) at the 1,000–2,000-cell stage; two examples (**c–d** and **e–i**) are shown. **c, d**, Note ectopic *ntl* staining (blue) at a distance from the *sqt*-expressing cells (brown) ($n = 7/18$). **e–g**, Images of the embryo in **h** and **i** after transplantation; **e** was taken shortly after transplantation; **f** and **g** were taken 1 h (**f**) and 2 h (**g**) after **e**. **h, i**, Transplanted cells (arrow in **g**) stained with *ntl* probe (arrow); **i**, High-magnification view of the *ntl*-expressing cells in **h** ($n = 6/12$).

MZ*oep* mutants and then transplanted several wild-type cells at a distance to the *sqt*-expressing cells. We also followed the movement of both *sqt*-expressing mutant cells and transplanted wild-type cells to ensure that they were not in close proximity between the time of transplantation and fixation (Fig. 4e–g).

Owing to the lack of *Oep* function in *sqt*-expressing and neighbouring cells, Sqt cannot activate the Nodal signalling pathway in these cells and thus no relay signal can be produced. If Sqt acts by a relay mechanism, the transplanted wild-type cells would not be able to express *ntl*. But if Sqt activates gene expression directly at a distance, wild-type cells should be able to receive the Sqt signal and express *ntl*. Consistent with the latter possibility, we found that wild-type cells in MZ*oep* mutant embryos can express *ntl* in response to a distant source of Sqt. This result suggests that Sqt activity produced in unresponsive (*oep* mutant) cells can traverse a field of unresponsive (*oep* mutant) cells to elicit an effect in distant responsive (wild-type) cells. These results show that the long-range effect of Sqt is direct, requiring no relay signal (Fig. 4c–i).

The idea of locally secreted signals that act as morphogens and directly pattern a tissue by eliciting distant responses has been an attractive mechanism to explain many developmental events^{1–10}. The potential of vertebrate mesoderm inducers to act as morphogens has been debated for a long time. Seemingly contradictory analyses of non-endogenous TGF- β signals have supported conflicting models. Elegant studies in *Xenopus* explants have revealed the potential of Activin to act as a long-range morphogen^{8–10}. In contrast, *in vivo* studies in zebrafish have suggested that Activin acts only locally and induces FGF as a relay signal¹². Similarly, Xnr2 and TGF- β 1 have been suggested to act only at short range^{9,11}. Our studies using the *in vivo* mesoderm inducers Cyc and Sqt now indicate that both short- and long-range mechanisms can operate during mesoderm induction and patterning. Notably, our results indicate that Sqt may act as a secreted morphogen: first, different levels of Sqt induce different downstream responses in a concentration- and distance-dependent way^{14,17–19} (Fig. 1); second, Sqt can act directly at a distance (Figs 2–4). These results indicate that Sqt acts as a secreted morphogen that does not require a relay mechanism. □

Methods

Single-cell injection

The lineage tracer fluorescein-dextran was co-injected with the RNA to mark the injected cell and its progeny. We synthesized *sqt*, *cyc* and *oep* sense, capped mRNAs as described¹⁸. RNA *in situ* hybridization and anti-fluorescein antibody staining were performed as described²².

Genotyping

After photography, embryos were washed twice in benzyl benzoate:benzyl alcohol (2:1), two or three times in 100% methanol and once in 100% ethanol. We extracted genomic DNA using the DNeasy Tissue Kit (Qiagen). After extraction, genomic DNA was precipitated and resuspended in 50 μ l H₂O. Five microlitres of the genomic DNA was used as a template in each polymerase chain reaction (PCR). Primers and conditions for genotyping *sqt*²³⁵ and *cyc*^{m294} have been described¹³.

Chimaera construction

Cyc–Sqt and Sqt–Cyc fusion proteins were constructed using PCR amplification. The junction sequence of Cyc–Sqt is RRGRNRHRT, in which the underlined sequence is the C terminus of the Cyc prodomain and the rest is the N terminus of the Sqt mature peptide. The junction sequence for Sqt–Cyc is RRHRRGPPVRR, in which the underlined sequence is the N terminus of the Cyc mature peptide and the rest is the C terminus of the Sqt prodomain.

Transplantation and triple labelling

Donor embryos were injected with biotin-dextran and rhodamine-dextran at the 1–4-cell stage. We carried out transplantation as described²². After *in situ* hybridization and detection of fluorescein-dextran, the biotin-dextran labelled donor cells were detected with the alkaline phosphatase substrate kit 1 (Vector Laboratories, Inc.).

Imaging

Fluorescent and Nomarski images were taken after mounting the embryo in 2.5% methyl cellulose. Images were acquired as described¹⁸.

Received 18 October 2000; accepted 26 February 2001

1. Turing, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B.* **237**, 37–72 (1952).
2. Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
3. Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a DPP morphogen gradient. *Cell* **85**, 357–368 (1996).
4. Lecuit, T. *et al.* Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. *Nature* **381**, 387–393 (1996).
5. Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a wingless morphogen gradient. *Cell* **87**, 833–844 (1996).
6. Neumann, C. J. & Cohen, S. M. Long-range action of Wingless organizes the dorsal–ventral axis of the *Drosophila* wing. *Development* **124**, 871–880 (1997).
7. Neumann, C. & Cohen, S. Morphogens and pattern formation. *BioEssays* **19**, 721–729 (1997).
8. Gurdon, J. B., Harger, P., Mitchell, A. & Lemaire, P. Activin signalling and response to a morphogen gradient. *Nature* **371**, 487–492 (1994).
9. Jones, C. M., Armes, N. & Smith, J. C. Signalling by TGF- β family members: short-range effects of Xnr-2 and BMP-4 contrast with the long-range effects of activin. *Curr. Biol.* **6**, 1468–1475 (1996).
10. McDowell, N., Zorn, A. M., Crease, D. J. & Gurdon, J. B. Activin has direct long-range signalling activity and can form a concentration gradient by diffusion. *Curr. Biol.* **7**, 671–681 (1997).
11. Reilly, K. M. & Melton, D. A. Short-range signaling by candidate morphogens of the TGF beta family and evidence for a relay mechanism of induction. *Cell* **86**, 743–754 (1996).
12. Rodaway, A. *et al.* Induction of the mesoderm in the zebrafish germ ring by yolk cell-derived TGF- β family signals and discrimination of mesoderm and endoderm by FGF. *Development* **126**, 3067–3078 (1999).
13. Feldman, B. *et al.* Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* **395**, 181–185 (1998).
14. Sampath, K. *et al.* Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. *Nature* **395**, 185–189 (1998).

15. Rebagliati, M. R., Toyama, R., Haffter, P. & Dawid, I. B. cyclops encodes a nodal-related factor involved in midline signaling. *Proc. Natl Acad. Sci. USA* **95**, 9932–7 (1998).
16. Schier, A. F. & Shen, M. M. Nodal signalling in vertebrate development. *Nature* **403**, 385–389 (2000).
17. Erter, C. E., Solnica-Krezel, L. & Wright, C. V. Zebrafish nodal-related 2 encodes an early mesodermal inducer signaling from the extraembryonic yolk syncytial layer. *Dev. Biol.* **204**, 361–372 (1998).
18. Gritsman, K., Talbot, W. S. & Schier, A. F. Nodal signaling patterns the organizer. *Development* **127**, 921–932 (2000).
19. Thisse, B., Wright, C. V. & Thisse, C. Activin- and Nodal-related factors control antero-posterior patterning of the zebrafish embryo. *Nature* **403**, 425–428 (2000).
20. Jones, C. M., Kuehn, M. R., Hogan, B. L., Smith, J. C. & Wright, C. V. Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. *Development* **121**, 3651–3662 (1995).
21. Agius, E., Oelgeschlager, M., Wessely, O., Kemp, C. & De Robertis, E. M. Endodermal Nodal-related signals and mesoderm induction in *Xenopus*. *Development* **127**, 1173–1183 (2000).
22. Gritsman, K. *et al.* The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. *Cell* **97**, 121–132 (1999).
23. Vogel, A. M. & Gerster, T. Promoter activity of the zebrafish bhlh1 retroelement requires an intact activin signaling pathway. *Mech. Dev.* **85**, 133–146 (1999).
24. Ruppert, R., Hoffmann, E. & Sebald, W. Human bone morphogenetic protein 2 contains a heparin-binding site which modifies its biological activity. *Eur. J. Biochem.* **237**, 295–302 (1996).
25. Pfeiffer, S., Alexandre, C., Calleja, M. & Vincent, J. P. The progeny of wingless-expressing cells deliver the signal at a distance in *Drosophila* embryos. *Curr. Biol.* **10**, 321–324 (2000).
26. Schier, A. F., Neuhauss, S. C., Helde, K. A., Talbot, W. S. & Driever, W. The one-eyed pinhead gene functions in mesoderm and endoderm formation in zebrafish and interacts with no tail. *Development* **124**, 327–342 (1997).
27. Zhang, J., Talbot, W. S. & Schier, A. F. Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation. *Cell* **92**, 241–251 (1998).
28. Strähle, U. *et al.* one-eyed pinhead is required for development of the ventral midline of the zebrafish (*Danio rerio*) neural tube. *Genes Funct.* **1**, 131–148 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank members of the Schier and Yelon laboratories for discussions; R. Burdine, A. Carmany-Rampey, K. Joubin, R. Lehmann, G. Struhl, W. Talbot and D. Yelon for comments on the manuscript; S. Zimmerman, R. Feeney and T. Bruno for fish care. Y.C. is the Rebecca Ridley Kryo Fellow of the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation. A.F.S. is a Scholar of the McKnight Endowment Fund for Neuroscience and the Irma T. Hirsch Trust, and is supported by grants from the NIH.

Correspondence and requests for materials should be addressed to A.F.S. (e-mail: schier@saturn.med.nyu.edu).

MOR1 is essential for organizing cortical microtubules in plants

Angela T. Whittington, Oliver Vugrek, Ke Jun Wei, Nortrud G. Hasenbein, Keiko Sugimoto*, Madeleine C. Rashbrooke & Geoffrey O. Wasteneys

Plant Cell Biology Group, Research School of Biological Sciences, The Australian National University, GPO Box 475, Canberra, ACT 2601, Australia

Microtubules orchestrate cell division and morphogenesis, but how they disassemble and reappear at different subcellular locations is unknown. Microtubule organizing centres are thought to have an important role, but in higher plants microtubules assemble in ordered configurations even though microtubule organizing centres are inconspicuous or absent. Plant cells generate highly organized microtubule arrays that coordinate mitosis, cytokinesis and expansion. Inhibiting microtubule assembly prevents chromosome separation¹, blocks cell division² and impairs growth polarity³. Microtubules are essential for the formation of cell walls, through an array of plasma-membrane-associated cortical microtubules whose control mechanisms are unknown. Using a genetic strategy to identify microtubule organizing factors in *Arabidopsis thaliana*, we isolated temperature-sensitive

mutant alleles of the *MICROTUBULE ORGANIZATION 1* (*MOR1*) gene. Here we show that MOR1 is the plant version of an ancient family of microtubule-associated proteins⁴. Point mutations that substitute single amino-acid residues in an amino-terminal HEAT repeat impart reversible temperature-dependent cortical microtubule disruption, showing that MOR1 is essential for cortical microtubule organization.

In most plant cells that display diffuse rather than tip growth, microtubules localize to the cortical cytoplasm perpendicular to the major axis of expansion. Microtubules and cellulose microfibrils often have similar orientation patterns in elongating cells⁵ and it is generally accepted, but not proven, that cortical microtubules control the alignment of cellulose microfibrils⁶. Identifying the factors that organize microtubule arrays at the periphery of plant cells is a necessary step towards understanding the mechanisms that underlie wall deposition and, hence, plant morphogenesis.

To identify factors regulating cortical microtubule organization in plant cells, we used immunofluorescence microscopy to screen chemically mutagenized seedlings of *A. thaliana* for aberrant microtubule patterns. One mutant locus, *mor1*, causes temperature-sensitive cortical microtubule shortening and disorganization (Fig. 1) and consequent morphological defects. We used ecotype-specific markers to identify MOR1 as a gene of around 14 kilobases (kb) that encodes a protein with a predicted relative molecular mass of 217,000 (*M_r* 217K) (Fig. 2a) that has significant deduced amino-acid sequence similarity to human TOGp⁷, *Xenopus* MAP215

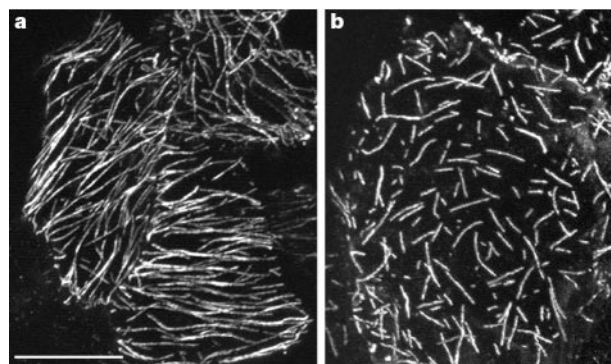


Figure 1 Mutations in the *MOR1* gene cause temperature-dependent microtubule disruption. We used anti-tubulin immunofluorescence to label cortical microtubules in epidermal cells of the first true leaf of 21-day-old seedlings after incubating seedlings at 29 °C for 2 h before fixation. **a**, Wild type. **b**, *mor1-1* homozygote. Scale bar, 25 μm.

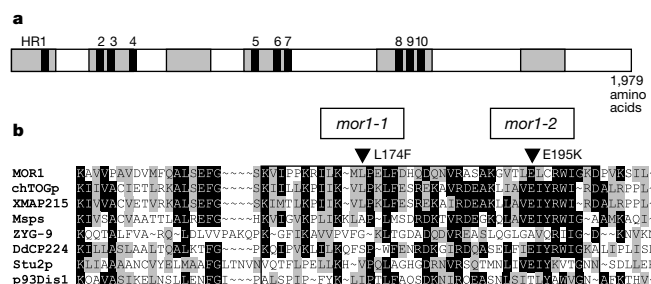


Figure 2 MOR1 protein structure. **a**, Schematic representation with shaded boxes indicating conserved domains between MOR1, TOGp, XMAP215, MSPs and DdCP224. The shorter ZYG-9 and yeast proteins share the first five and first four regions of homology, respectively. Black stripes show putative HEAT repeats (HR) in relation to these conserved domains. **b**, Deduced amino-acid sequence comparison of MOR1's HEAT repeat-1 with equivalent repeats in homologues. Black-shaded residues indicate identity, grey-shaded residues indicate similarity (40% threshold), outlined box indicates HEAT repeat. Mutations altering MOR1 amino-acid residues are indicated by arrowheads.

* Present address: John Innes Centre, Department of Cell and Developmental Biology, Norwich NR4 7UH, UK.

(ref. 4), *Drosophila* MSPS⁸ and *Dictyostelium* DdCP224 (ref. 9), and more limited homology to yeast STU2 (ref. 10) and Dis1 (ref. 11) and the ZYG-9 protein of *Caenorhabditis elegans*¹², all members of a family of microtubule-associated proteins⁴. Polymerase chain reaction with reverse transcription (RT-PCR) generated roughly 6-kb complementary DNA transcripts from total RNA of wild-type Columbia and *mor1* mutant seedlings. Comparison of the genomic and cDNA sequences shows that *MOR1* comprises 53 exons (see Supplementary Information).

Sequence analysis¹³ identified at least ten putative HEAT repeats in *MOR1*, which represent sites of likely protein–protein interaction¹⁴ (Fig. 2a). The two identified *mor1* mutant alleles each have single amino-acid substitutions in the HEAT repeat nearest the N terminus. In the *mor1-1* mutant allele, a cytosine-to-thymine base change substitutes the amino acid leucine at position 174 with phenylalanine, whereas in *mor1-2*, a guanine-to-adenine change replaces the glutamic acid at position 195 with lysine (Fig. 2b). In this region, deduced amino-acid sequence conservation is high and a HEAT repeat is predicted for all known *MOR1* homologues (Fig. 2b). This conserved motif could be a microtubule-binding site, as it falls within the region shown to be essential for microtubule binding in some of *MOR1*'s homologues^{9,15}. HEAT repeats, however, can also confer plasma membrane localization¹⁶, so it remains possible that the HEAT repeat identified by the *mor1* mutations mediates localization of cortical microtubules.

Immunofluorescence microscopy indicates that the *mor1* mutations cause cortical microtubules to shorten and lose ordered alignment at temperatures above 28 °C (Fig. 1). Short microtubules are also observed in *zyg9* mutants¹² of *C. elegans* and in *Xenopus* egg extracts depleted of XMAP215 (ref. 4). To characterize the mutant

phenotype in living cells, we expressed a green fluorescent protein (GFP)–tubulin fusion protein¹⁷ in *mor1* mutants and documented microtubule disruption and recovery. At 21 °C, *mor1* microtubule patterns in day 5 hypocotyl epidermal cells (Fig. 3a) were indistinguishable from those in equivalent wild-type cells, but subtle changes were observed almost immediately after shifting to 29 °C. Cortical microtubule arrays thereafter became progressively disrupted over a 90-min period (Fig. 3b–g) compared with wild-type cells (Fig. 3j). Microtubules with non-transverse orientation became more common and patchy diffuse fluorescence developed, suggesting dissociation of tubulin subunits from intact microtubules. Elliptical clear zones 4–6 µm in length, surrounded by bright fluorescence (Fig. 3g, inset), were observed after around 30 min at 29 °C and became increasingly abundant. Eventually, the microtubules became relatively short, sparse and misaligned, but they never completely disassembled (Fig. 3g).

Recovery of normal microtubule organization after returning to the permissive temperature was rapid, suggesting that a temperature-induced change in protein structure caused the observed loss of structural organization at the restrictive temperature. Within a few minutes of reaching the permissive temperature, background fluorescence diminished, the elliptical clear zones disappeared, microtubules lengthened and parallel order was steadily restored (Fig. 3h, i). Microtubule arrays did not always recover transverse orientation but this was consistent with the developmental progression from transverse to oblique microtubule orientation over the 4-h treatment period¹⁸. At the permissive temperature, cortical microtubules have a better capacity to assemble, consistent with the promotion of microtubule plus-end assembly in *MOR1* homologues¹⁹.

We identified no constitutive mutant alleles of *MOR1* and suspect that these would cause seedling lethality. At the permissive

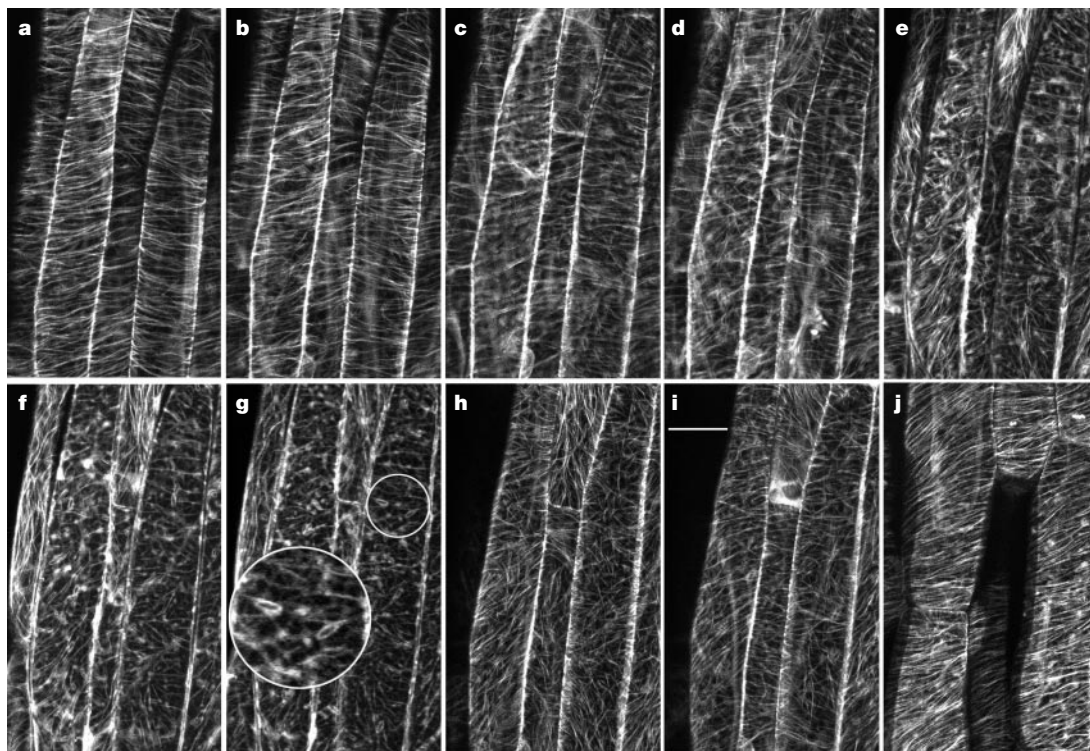


Figure 3 Temperature-dependent microtubule disruption in *mor1* mutants is reversible. Cortical microtubule arrays in elongating hypocotyl epidermal cells were rendered fluorescent by expression of the GFP–TUA6 fusion protein, and imaged by confocal microscopy. **a–i**, A time-lapse series from the same area. **a**, Microtubules are mostly transversely aligned at 21 °C but gradually become sparse and misaligned when the temperature is increased to 29 °C (**b–g**). **b**, 0 min; **c**, 16 min; **d**, 30 min; **e**, 60 min;

f, 75 min; **g**, 90 min, all at 29 °C. Enlarged circle in **g** shows magnified image of elliptical clear zones. After 2 h at the restrictive temperature, restoration of the permissive temperature (21 °C) results in rapid change in the appearance of the microtubules (**h**, **i**). **h**, 2 h at 29 °C, 7 min at 21 °C; **i**, 2 h at 29 °C, 48 min at 21 °C. **j**, In wild-type cells treated at 29 °C for 2 h, there is no sign of microtubule disruption. Scale bar, 10 µm.

temperature, *mor1* homozygotes and wild-type plants cannot be distinguished, but *mor1* seedlings germinated at the restrictive temperature are extremely squat, do not develop flowers (Fig. 4a) and eventually die. Closer examination of plants at the permissive (Fig. 4c) and restrictive temperatures revealed numerous morphological defects including left-handed twisting of organs (Fig. 4d), isotropic cell expansion (Fig. 4e) and impaired root hair polarity (Fig. 4f). At any stage of development the *mor1* mutants show altered morphology and function at the restrictive temperature. Shifting to the restrictive temperature after the onset of flowering, for example, alters floral morphology and abolishes fertility. MOR1 is therefore required throughout development and in all organs. RT-PCR analysis confirmed expression of MOR1 in roots, cotyledons, rosette leaves, stems, open flowers

and green siliques at permissive and restrictive temperatures (data not shown). Interestingly, rosette leaves, although severely deformed, do develop at the restrictive temperature (Fig. 4a), showing that cell division continues. Restoring the permissive temperature, even after weeks at the restrictive temperature, initiates recovery (Fig. 4b). Plants bloom and produce viable seed that gives rise to progeny with normal ploidy levels, indicating that the integrity of the apical meristem is preserved.

The *mor1* mutations did not affect mitotic and cytokinetic microtubule arrays. At the restrictive temperature, preprophase bands, mitotic spindles and phragmoplasts in *mor1* mutants were indistinguishable in appearance from those of wild-type cells, although cortical arrays in adjacent interphase cells were disrupted (data not shown). There was also no sign of defective cell patterns or reduced cell numbers in the tissues of mutant plants, despite obvious increases in the radial dimensions of expanding cells (Fig. 4g–j). Although this indicates that MOR1 may not participate in organizing mitotic and cytokinetic microtubules, all of the known MOR1 homologues associate with mitotic microtubule arrays^{8–12,20}. Phosphorylation of *Xenopus* XMAP215 by the mitotic kinase CDK1 reduces XMAP215's ability to polymerize microtubules²¹ and XMAP215 isolated from mitotic cells has reduced microtubule binding affinity²². A similar cell-cycle-specific modulation could account for the *mor1* mutant's effect on interphase cortical but not mitotic microtubules. Whether MOR1 is regulated by a CDK1 homologue remains unclear because we have found no matches in the MOR1 sequence for the CDK1 consensus phosphorylation site²³. We also considered the possibility that a MOR1 orthologue could specialize in organizing mitotic and cytokinetic arrays. Southern blotting and genome sequence analysis, however, indicate that *MOR1* is a single copy gene; two apparent pseudogenes lack functional amino- and carboxy-terminal regions.

Identifying microtubule-associated proteins in plants has been a slow process. The fact that plant microtubule arrays differ in complexity from those of animal cells predicts that some plant MAPs will be altogether different from animal MAPs^{24,25}. Indeed, the MAP65 class of plant MAPs has no known animal homologues²⁶. MOR1's homology to the TOGP–XMAP215 class of MAPs, however, proves that at least one structural MAP is conserved in animal and plant cells. Our mutational strategy shows that MOR1 is essential for organizing the dispersed cortical microtubule array that regulates expansion properties of plant cells. This discovery has important implications not only for understanding the specialized features and functions of microtubule arrays in plant cells but also for identifying the mechanisms by which the TOGP–XMAP215 family of proteins carry out their essential microtubule-organizing function. □

Methods

Mutagenesis and mutant screening

We used anti-tubulin immunofluorescence of cotyledon or first leaf epidermal cells²⁷ to identify mutants with temperature-dependent microtubule disorganization. Out of 10,000 seed lines treated with ethyl methane sulphonate, two plants were identified that had short and misaligned cortical microtubules after 2 h incubation at 29 °C.

Microtubule phenotype characterization

To examine microtubules in living cells, we crossed isogenic *mor1* mutant lines with a transgenic line constitutively expressing a fusion protein of GFP and *A. thaliana* α -tubulin 6, driven by the 35S promoter of cauliflower mosaic virus¹⁷. F₂ segregants expressing GFP and F₃ progeny from *mor1* homozygous F₂ plants were used for subsequent experiments. Intact 5-day-old seedlings were placed in incubation chambers so that the hypocotyl was against the coverslip. Using confocal laser scanning microscopy, we examined cells nearing the completion of elongation that still had predominantly transverse cortical microtubules. A temperature-controlled stage regulated the rapid shifts between 21 °C and 29 °C.

Mapping the *MOR1* locus

mor1-1 was crossed with the Landsberg *erecta* (Ler) ecotype and DNA extracted from

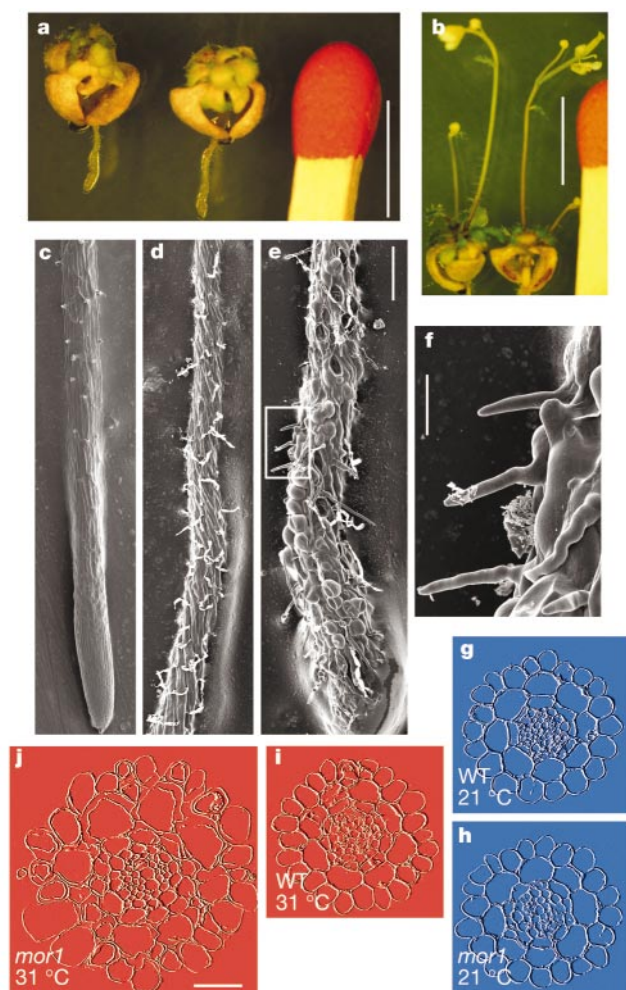


Figure 4 Morphological consequences of *mor1*'s microtubule disruption. **a**, Cultured for 4 weeks at 31 °C, then 1 week at 21 °C, *mor1-1* (left) and *mor1-2* (right) plants are severely stunted, show radially swollen and short organs and do not produce flowers. Scale bar, 5 mm. **b**, After culture at 31 °C for 2 weeks followed by 3 weeks at 21 °C, latent ability to bloom is triggered in both *mor1-1* (left) and *mor1-2* (right) mutants. Scale bar, 5 mm. Scanning electron micrographs of cryo-preserved *mor1-1* mutant root tips (**c–f**) show morphological defects in closer detail. **c**, Normal growth at 21 °C. **d**, After 6 days at 31 °C, anti-clockwise twisting and leftward bending are evident in upper portion of root. **e**, In lower, younger part of same root, radial swelling is greatest. Scale bar for **c–e**, 100 μ m. **f**, Higher magnification from boxed area in **e**, showing crooked root hairs. Scale bar, 50 μ m. Root transverse sections from elongation zones of 7-day-old roots (**g–j**) show that tissue patterns and cell numbers are not altered in *mor1* mutants. **g**, Wild type (WT), 21 °C. **h**, *mor1-1*, 21 °C. **i**, Wild type, 31 °C for 48 h. **j**, *mor1-1*, 31 °C for 48 h. Scale bar for **g–j**, 50 μ m.

homozygous F₂ or F₃ *mor1-1* segregants was used to map the *MOR1* locus using CAPS (cleaved amplified polymorphic sequence) markers²⁸ to distinguish parental (Columbia) from *Ler* DNA sequence. Tight linkage was established to two CAPS markers flanking (m323 and C43/C44), and to one CAPS marker (Ubi) located within, bacterial artificial chromosome (BAC) T20F21 (Arabidopsis Biological Resource Centre, GenBank accession no. AC006068) at ~67.19 cM on chromosome 2. Homology searches of predicted genes on BAC T20F21 identified the most likely candidate gene as AC006068.3. Genomic and RT-PCR product sequence analysis identified point mutations in both mutant alleles. Transformation of *mor1-1* homozygotes with a subclone of BAC T20F21 containing *MOR1* restored the wild-type phenotype at the restrictive temperature, confirming the identity of the candidate gene. The wild type sequence from RT-PCR analysis is deposited in GenBank (accession no. AF367246).

Sequence analysis

Manual adjustment of the ClustalW multiple alignment used the BioEdit²⁹ program with reference to previously published pairwise alignments^{8–10,12}.

Received 23 February; accepted 17 April 2001.

1. Baskin, T. I. & Cande, W. Z. The structure and function of the mitotic spindle in flowering plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41**, 277–315 (1990).
2. Mayer, U., Herzog, U., Berger, F., Inze, D. & Jurgens, G. Mutations in the *pilz* group genes disrupt the microtubule cytoskeleton and uncouple cell cycle progression from cell division in Arabidopsis embryo and endosperm. *Eur. J. Cell Biol.* **78**, 100–108 (1999).
3. Baskin, T. I., Wilson, J. E., Cork, A. & Williamson, R. E. Morphology and microtubule organization in Arabidopsis roots exposed to oryzalin or taxol. *Plant Cell Physiol.* **35**, 935–942 (1994).
4. Tournebise, R. *et al.* Control of microtubule dynamics by the antagonistic activities of XMAP215 and XKCM1 in *Xenopus* egg extracts. *Nature Cell Biol.* **2**, 13–19 (2000).
5. Sugimoto, K., Williamson, R. E. & Wasteneys, G. O. New techniques enable comparative analysis of microtubule orientation, wall texture and growth rate in intact roots of Arabidopsis thaliana. *Plant Physiol.* **124**, 1493–1506 (2000).
6. Wasteneys, G. O. The cytoskeleton and growth polarity. *Curr. Opin. Plant Biol.* **3**, 503–511 (2000).
7. Charrasse, S. *et al.* The TOGp protein is a new human microtubule-associated protein homologous to the *Xenopus* XMAP215. *J. Cell Sci.* **111**, 1371–1383 (1998).
8. Cullen, C. F., Deak, P., Glover, D. M. & Ohkura, H. Mini spindles: A gene encoding a conserved microtubule-associated protein required for the integrity of the mitotic spindle in *Drosophila*. *J. Cell Biol.* **146**, 1005–1018 (1999).
9. Graf, R., Daudeker, C. & Schliwa, M. Dictyostelium DdCP224 is a microtubule-associated protein and a permanent centrosomal resident involved in centrosome duplication. *J. Cell Sci.* **113**, 1747–1758 (2000).
10. Wang, P. J. & Huffaker, T. C. Stu2p: A microtubule-binding protein that is an essential component of the yeast spindle pole body. *J. Cell Biol.* **139**, 1271–1280 (1997).
11. Nabeshima, K. *et al.* p93dis1, which is required for sister chromatid separation, is a novel microtubule and spindle pole body-associated protein phosphorylated at the Cdc2 target sites. *Genes Dev.* **9**, 1572–1585 (1995).
12. Matthews, L. R., Carter, P., Thierry-mieg, D. & Kempheus, K. Zyg-9, a *Caenorhabditis elegans* protein required for microtubule organization and function, is a component of meiotic and mitotic spindle poles. *J. Cell Biol.* **141**, 1159–1168 (1998).
13. Andrade, M. A., Ponting, C. P., Gibson, T. J. & Bork, P. Homology-based method for identification of protein repeats using statistical significance estimates. *J. Mol. Biol.* **298**, 521–537 (2000).
14. Andrade, M. A. & Bork, P. Heat repeats in the Huntingtons disease protein. *Nature Genet.* **11**, 115–116 (1995).
15. Spittle, C., Charrasse, S., Larroque, C. & Cassimeris, L. The interaction of TOGp with microtubules and tubulin. *J. Biol. Chem.* **275**, 20748–20753 (2000).
16. Kunz, J., Schneider, U., Howald, I., Schmidt, A. & Hall, M. N. HEAT repeats mediate plasma membrane localization of Tor2p in yeast. *J. Biol. Chem.* **275**, 37011–37020 (2000).
17. Ueda, K., Matsuyama, T. & Hashimoto, T. Visualization of microtubules in living cells of transgenic Arabidopsis thaliana. *Protoplasma* **206**, 201–206 (1999).
18. Ueda, K. & Matsuyama, T. Rearrangement of cortical microtubules from transverse to oblique or longitudinal in living cells of transgenic Arabidopsis thaliana. *Protoplasma* **213**, 28–38 (2000).
19. Gard, D. L. & Kirchner, M. W. A microtubule-associated protein from *Xenopus* eggs that specifically promotes assembly at the plus-end. *J. Cell Biol.* **105**, 2203–2215 (1987).
20. Charrasse, S., Lorca, T., Doree, M. & Larroque, C. The *Xenopus* XMAP215 and its human homologue TOG proteins interact with cyclin B1 to target p34cdc2 to microtubules during mitosis. *Exp. Cell Res.* **254**, 249–256 (2000).
21. Vasquez, R. J., Gard, D. L. & Cassimeris, L. Phosphorylation by CDK1 regulates XMAP215 function in vitro. *Cell. Motil. Cytoskeleton* **43**, 310–321 (1999).
22. Andersen, S. S. *Xenopus* interphase and mitotic microtubule-associated proteins differentially suppress microtubule dynamics in vitro. *Cell. Motil. Cytoskeleton* **41**, 202–213 (1998).
23. Kennelly, P. J. & Krebs, E. G. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266**, 15555–15558 (1991).
24. Wick, S. Plant microtubules meet their MAPs and mimics. *Nature Cell Biol.* **2**, E204–E206 (2000).
25. Lloyd, C. W. & Hussey, P. J. Microtubule-associated proteins in plants—why we need a MAP. *Nature Rev. Mol. Cell Biol.* **2**, 40–47 (2001).
26. Smertenko, A. *et al.* A new class of microtubule-associated proteins in plants. *Nature Cell Biol.* **2**, 750–753 (2000).
27. Wasteneys, G. O., Willingale-Theune, J. & Menzel, D. Freeze shattering: a simple and effective method for permeabilizing higher plant cell walls. *J. Microsc.* **188**, 51–61 (1997).
28. Konieczny, A. & Ausubel, F. M. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403–410 (1993).
29. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

R. E. Williamson provided support from the inception of this project and supplied EMS-mutagenized seed stock. The Arabidopsis Biological Resource Centre, Ohio State University, supplied BACs. We thank T. Hashimoto and K. Ueda for the GFP::TUA6 line, M. Luo and A. Chaudhury for the binary clone used in complementation, H. Kazama for helpful advice, R. Heady and A. Knox for microscopy assistance, and J. B. Gibson, B. E. S. Gunning and A. R. Hardham for comments on the manuscript. An Australian Research Council QEII Fellowship to G.O.W. supported this work. Coauthor contributions include: mutant identification and genetic analysis (G.O.W. and N.G.H.), gene mapping (O.V., K.-J.W., A.T.W. & G.O.W.), cloning and sequence analysis (A.T.W., K.-J.W., M.C.R. & G.O.W.), microtubule dynamics (G.O.W.) and morphological analysis (K.S., N.G.H. & G.O.W.).

Correspondence and requests for materials should be addressed to G.O.W. (e-mail: geoffw@rsbs.anu.edu.au). Sequences are deposited in GenBank under accession no. AF367246.

Metastasis suppressor gene *KiSS-1* encodes peptide ligand of a G-protein-coupled receptor

Tetsuya Ohtaki*, Yasushi Shintani*, Susumu Honda*, Hirokazu Matsumoto*, Akira Hori*, Kimiko Kanehashi*, Yasuko Terao*, Satoshi Kumano*, Yoshihiro Takatsu*, Yasushi Masuda*, Yoshihiro Ishibashi*, Takuya Watanabe*, Mari Asada*, Takao Yamada*, Masato Suenaga*, Chieko Kitada*, Satoshi Usuki†, Tsutomu Kurokawa*, Haruo Onda*, Osamu Nishimura* & Masahiko Fujino*

* Pharmaceutical Discovery Research Division, Takeda Chemical Industries Ltd, Wada 10, Tsukuba, Ibaraki 300-4293, Japan

† Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8575, Japan

Metastasis is a major cause of death in cancer patients and involves a multistep process including detachment of cancer cells from a primary cancer, invasion of surrounding tissue, spread through circulation, re-invasion and proliferation in distant organs. *KiSS-1* is a human metastasis suppressor gene¹, that suppresses metastases of human melanomas² and breast carcinomas³ without affecting tumorigenicity. However, its gene product and functional mechanisms have not been elucidated. Here we show that *KiSS-1* (refs 1, 4) encodes a carboxy-terminally amidated peptide with 54 amino-acid residues, which we have isolated from human placenta as the endogenous ligand of an orphan G-protein-coupled receptor (hOT7T175) and have named 'metastin'. Metastin inhibits chemotaxis and invasion of hOT7T175-transfected CHO cells *in vitro* and attenuates pulmonary metastasis of hOT7T175-transfected B16-BL6 melanomas *in vivo*. The results suggest possible mechanisms of action for *KiSS-1* and a potential new therapeutic approach.

During a search for novel G-protein-coupled receptors (GPCRs) using degenerate polymerase chain reaction (PCR) strategy we found a rat orphan receptor (rOT7T175) that was nearly identical to GPR54 (ref. 5). To identify the endogenous ligand of OT7T175, we established a stable CHO cell line (CHO/h175) expressing the human counterpart hOT7T175. Although hOT7T175 had 39.2% amino-acid identity to human galanin receptor GALR2 (ref. 6), the cells did not show any response to a panel of known peptides, including galanin and galanin-like peptide⁷. On the other hand, human placental extract induced a robust increase in intracellular calcium ion concentration ([Ca²⁺]_i) in CHO/h175 cells. After several steps of high-performance liquid chromatography purification for this activity, we isolated about 6 µg of pure peptide from 740 g of chorionic tissue (Fig. 1a, b).

The amino-terminal sequence obtained for the isolated peptide, GTSLSPPPSSGSRQQPGLXA (X not identified), was identical to the partial amino-acid sequence (Gly 68 to Ala 88) of the *KiSS-1* gene product^{1,4}. Although the *KiSS-1* gene product has been proposed to be an intracellular signalling molecule such as an SH3 ligand or a substrate for protein kinase^{1,4}, it has an obvious secretory signal sequence⁸ and two potential processing sites⁹ (Fig. 1d). Processing at these sites followed by amide transfer⁹ produces the *KiSS-1* gene-product (68–121) amide. Mass spectrometric analysis revealed that the *m/z* value of the isolated peptide (5,857.2) was very close to the theoretical value of the *KiSS-1* gene-product (68–121) amide (5,858.5). The *m/z* values of V8 protease-digested peptides (4,006.4 and 1,869.8) were in accordance with the theoretical value of the *KiSS-1* gene-product (68–106) amide

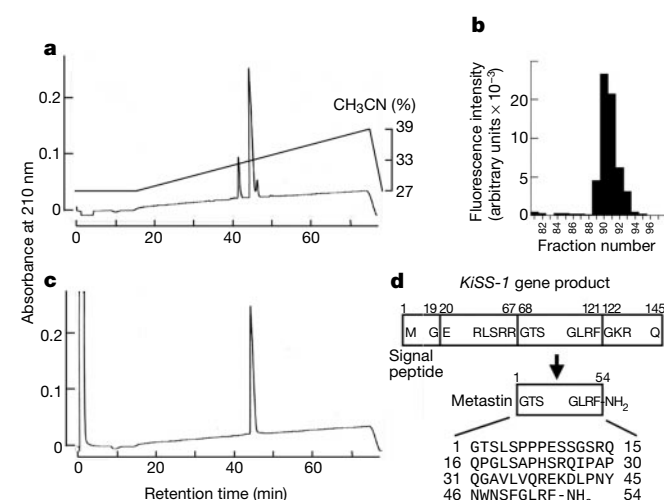


Figure 1 Purification of metastin. **a**, The elution profile of placental metastin at the final Super-ODS step. **b**, Each 0.5-min fraction of the eluate (from **a**) was collected and analysed by FLIPR assay with CHO/h175 cells. **c**, The elution profile of recombinant metastin under the same conditions as **a**. **d**, Hypothetical processing pathway of metastin in neuroendocrine tissues. After removal of the signal peptide Met 1 to Gly 19 (predicted by the SignalP server⁸), metastin precursor was cleaved at Arg 67 and at Arg 124 by a prohormone convertase. The C-terminal extended Arg 124 and Lys 123 were sequentially removed by carboxypeptidase, and Phe 121 was transamidated with Gly 122.

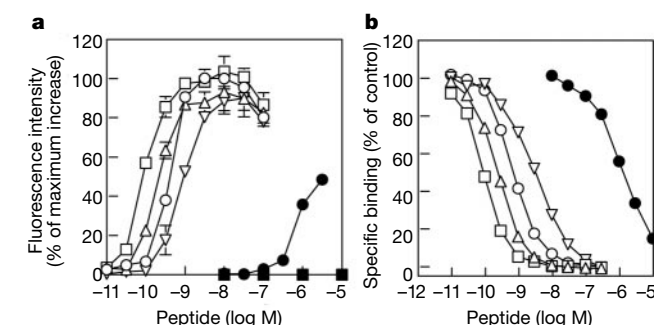


Figure 2 Pharmacological characterization. **a**, Response of CHO/h175 cells to metastin (open circle), metastin (40–54) (open uptriangle), (45–54) (open square), (46–54) (open down triangle) and metastin free form (filled circle) was measured by FLIPR assay. Response of CHO/mock cells to metastin (filled square). **b**, Competitive binding analysis using 100 pM [¹²⁵I]-Tyr⁴⁵ metastin (40–54). Symbols are the same as in **a**. The *K_i* values¹⁷ were: metastin, 0.34 nM; metastin (40–54), 0.10 nM; (45–54), 0.042 nM; (46–54), 1.5 nM; and free form, 640 nM.

(4,006.4) and that of the (107–121) amide (1,869.9), respectively. The isolated peptide was thus identified as the *KiSS-1* gene-product (68–121) amide and designated metastin.

Metastin and its derivative peptides were synthesized chemically or by a recombinant DNA method¹⁰ (Fig. 1c). Metastin potently and specifically induced an increase in [Ca²⁺]_i in CHO/h175 cells (Fig. 2a). The C-terminally free (un-amidated) form of metastin showed only very weak activity. N-terminally truncated peptides, metastin (40–54) and (45–54) were three to ten times more active than metastin, whereas metastin (46–54) was less active. Thus, the C-terminally amidated sequence from Tyr 45 to Phe 54 is mostly involved in receptor interaction. Saturation binding analysis showed that [¹²⁵I]-Tyr⁴⁵metastin (40–54) binds to the membranes of CHO/h175 cells in a saturable manner with a *K_d* (dissociation constant) of 95 pM. The receptor binding affinity of each peptide was obtained by competitive binding analysis (Fig. 2b). The affinity of metastin (inhibition constant (*K_i*) = 0.34 nM) was high enough to show it to be a cognate ligand for hOT7T175. Although truncated peptides had higher receptor binding affinity than metastin, these peptides were not endogenous. The N-terminal portion of metastin is not essential for receptor binding, but it may be involved in another biological process such as stabilization and protection from proteolytic digestion.

We further studied whether metastin can suppress tumour metastasis. First, we tested the effect of peptide on cell motility, a feature that is considered to be crucial for invasion of cancer cells. The addition of metastin to the upper microchemotaxis chamber inhibited the chemotaxis of CHO/h175 cells towards fetal calf serum (FCS), but did not affect that of the CHO mock transfectants (CHO/mock) (Fig. 3a). The C-terminally free form of metastin had 10,000-fold less activity, which was in good agreement with the

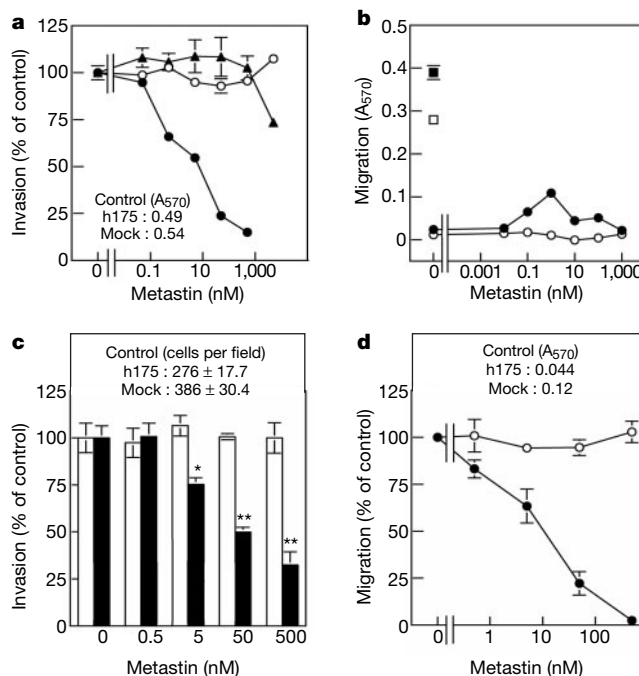


Figure 3 Cell migration and invasion. **a**, Migration of CHO/h175 (filled symbols) and CHO/mock cells (open symbols) towards FCS in the presence of metastin (circles) or its free form (triangles). **b**, Migration of CHO/h175 (filled symbols) and CHO/mock cells (open symbols) towards metastin (circles) or FCS (squares) for 4 h. **c**, Invasion of CHO/h175 (filled bars) and CHO/mock cells (open bars). **d**, Migration of B16-BL6/h175 (filled circles) and B16-BL6/mock melanomas (open circles) towards fibronectin. Values are mean $\pm$ s.e. (**a**, **b**, **d**) or mean $\pm$ s.d. (**c**). *, *P* < 0.05; **, *P* < 0.01 (Student's *t*-test).

pharmacological profile. Metastin by itself exerted weak chemotactic activity on CHO/h175 cells, but the maximum extent was only one-fourth of that induced by FCS and the activity disappeared at higher concentrations (10–1,000 nM) (Fig. 3b). In an *in vitro* invasion assay that tests migration through a Matrigel-coated porous filter, metastin inhibited FCS-induced invasion of CHO/h175 cells specifically with an IC₅₀ value of about 50 nM (Fig. 3c).

We transfected hOT7T175 to B16-BL6 mouse melanoma and obtained a stable high expression clone (B16-BL6/h175) that showed [Ca²⁺]_i response to metastin. Metastin inhibited the fibronectin-induced chemotaxis of B16-BL6/h175 melanomas but not of the mock-transfected melanomas (B16-BL6/mock) (Fig. 3d). Next, we found that metastin induced excessive formation of focal adhesions (vinculin-positive patches) and stress fibres in B16-BL6/h175 melanomas (Fig. 4a, b). The cells started to express focal adhesions 5 min after the addition of metastin, and most of the cells displayed focal adhesions densely after 30 min. Excessive stress fibres were formed after 30 min but not after 5 min. Furthermore, metastin induced the phosphorylation of focal adhesion kinase (FAK) and paxillin (Fig. 4c, d), which is believed to be essential for the formation of focal adhesions¹¹. The phosphorylation of FAK occurred rapidly and almost ceased after 10 min, whereas that of paxillin lasted longer. We hypothesize that metastin attenuates cellular motility by excessively inducing the adhesive phenotypes. The detailed signal transduction mechanisms should be studied in future. Sphingosine-1-phosphate is also known to inhibit the migration of B16 melanoma¹², which is mediated by a GPCR named Edg-5 and involves subsequent activation of Rho, FAK and paxillin, and overexpression of focal adhesions¹³.

In spontaneous pulmonary metastasis assays, B16-BL6/h175 melanomas were injected into the footpads of mice on day 0 (d0), tumours were surgically excised on d21, and metastasized tumour foci were counted macroscopically on d35. As the lung had only a small number of metastasized cells (microscopic inspection) on d21, administration of metastin with an osmotic mini-pump, limited to two weeks, was started on d18. This was also to mimic a potential clinical setting. When tumours were surgically removed on d21, there were not significant differences in tumour size of metastin- and vehicle-treated mice (Table 1). On d35, metastin-treated mice had a marked decrease in the number of metastasized tumour foci compared with mice that had received vehicle (Table 1). The number of microscopic metastases also decreased in the

Table 1 Metastasis of B16-BL6/h175 melanomas

Cell injection	Administration	Tumour size (mm ³)*	Tumour foci (no.)	n
B16-BL6/h175 (s.c.)	Metastin	900.8 ± 97.6†	13.0 ± 3.2†	5
	Vehicle	1,037.2 ± 170.5	37.8 ± 2.5	4
B13-BL6/mock (s.c.)	Metastin	1,062.9 ± 81.7†	31.2 ± 6.6†	10
	Vehicle	1,135.2 ± 99.6	26.4 ± 6.8	8
B16-BL6/h175 (i.v.)	Metastin		59.8 ± 9.8†	6
	Vehicle		42.7 ± 6.9	6

*Tumour size on day 21.

†Not significant compared with vehicle.

‡P < 0.01 compared with vehicle (Student's t-test). Values are mean ± s.e.

metastin-treated mice (see Supplementary Information). On the other hand, metastin did not affect metastasis of B16-BL6/mock melanomas (Table 1). Therefore, the observed effect on tumour metastasis is probably due to the direct action of metastin on tumour cells, although metastin may have also exerted some pharmacological effect on host mice. On the basis of the inhibitory activity on cell migration and invasion *in vitro*, metastin can be considered to have caused the inhibitory activity on the invasion of primary tumours and re-invasion of circulating tumour cells. Furthermore, metastin may have attenuated the microgrowth of metastasizing cells, as metastin was effective despite being administered relatively late (d18). By contrast, metastin failed to block the experimental metastasis (Table 1) and the growth of tumour mass *in vivo* (see Supplementary Information). However, these findings do not eliminate the possibility of metastin action at lower cell density. The precise mechanism of the antimetastatic activity remains to be elucidated.

We investigated the tissue distribution of *KiSS-1* and *hOT7T175* transcripts in normal and tumour tissues using a quantitative PCR with reverse transcription (RT-PCR) method. Very high *KiSS-1* gene expression was observed in placenta, with the next highest level in testis and moderate levels in pancreas, liver and small intestine (Fig. 5). For *hOT7T175* transcripts, the highest expression levels were found in pancreas and placenta, with moderate expression in spleen, the peripheral blood leukocyte (PBL), testis and lymph node. In the matched complementary DNA pairs that included cDNAs prepared from tumour and normal adjacent tissues of individual patients, tumour samples from four patients had *hOT7T175* expression levels as high as the highest levels observed in pancreas and placenta, whereas normal tissues had low to

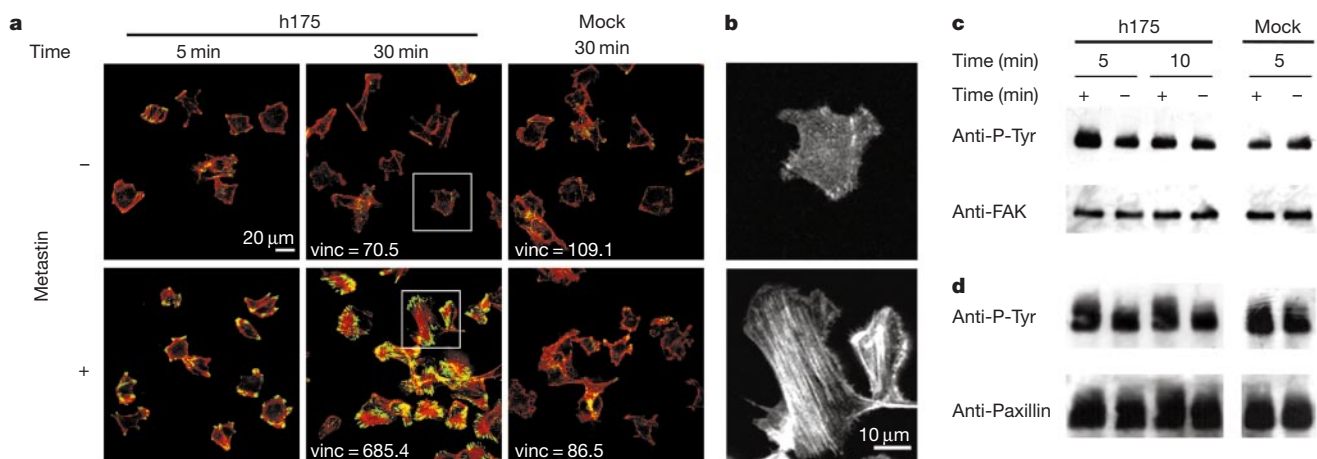


Figure 4 Formation of focal adhesions and stress fibres. B16-BL6/h175 (h175) or B16-BL6/mock melanomas (mock) were challenged with 10 μ M metastin (+) or vehicle (–) for indicated period. **a**, Staining with phalloidin (red) and anti-vinculin antibody (green). Vinculin-stained area per single cell is shown in average of arbitrary units (vinc). **b**, High

power enlargement of phalloidin-stained image from insets in **a**, **c**, **d**. Immunoprecipitate with anti-FAK (**c**) or anti-paxillin antibody (**d**) was analysed by western blotting with indicated antibody.

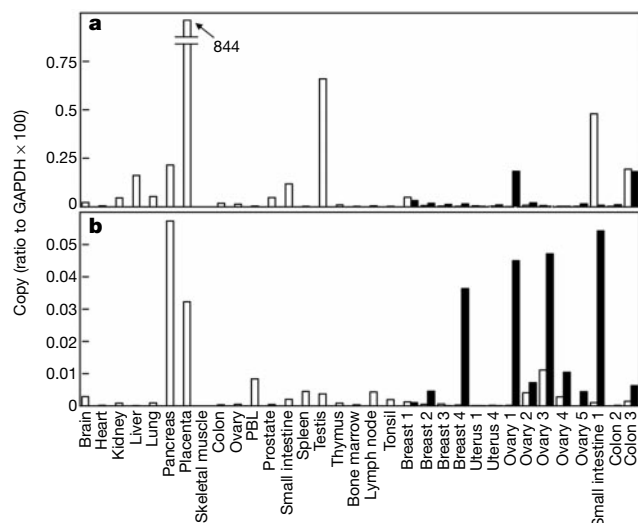


Figure 5 Quantitative RT-PCR analyses of *KiSS-1* and *hOT7T175* transcripts in human multiple tissue cDNA panels and matched cDNA pairs (Clontech). White bars, normal tissues; black bars, tumour tissues. **a**, *KiSS-1*; **b**, *hOT7T175*.

negligible levels of expression (Fig. 5). By contrast, high expression of *KiSS-1* gene was not found often in these tumours, but was found in normal tissues of two patients. From these observations, we infer that some human cancers express *hOT7T175* but are deficient in metastin. These cancer tissues are potential clinical targets of metastin. It is however unclear whether metastin is physiologically involved in the regulation of cell motility. The best characterized regulation is found in placenta, where the invasive behaviour of cytotrophoblasts is spatially and temporally regulated by metalloproteases, cytokines and growth factors¹⁴. Considering the abundance of metastin in human placenta, possible involvement of metastin in the regulation of cytotrophoblasts should be investigated in future studies.

The *KiSS-1* gene encodes metastin, the endogenous peptide ligand of the GPCR *hOT7T175*. Metastin inhibits the chemotaxis, invasion and metastasis of *hOT7T175*-expressing cells. We propose that the antimetastatic effect of the *KiSS-1* gene may be mediated by metastin and its receptor. It should be investigated whether C8161 melanoma and MDA-MB-435 breast carcinoma used in studies elsewhere^{1–3} express *hOT7T175* receptor, and whether *KiSS-1* gene-transfected tumour cells secrete metastin. Further studies on the physiological roles and clinical potential of this peptide shall follow.

Methods

Cloning of *hOT7T175* cDNA

A large part of *hOT7T175* sequence was found in the KAZUSA Genomic DNA Database (VTS35556). The full-length *hOT7T175* cDNA was cloned from a brain cDNA library using a GeneTrapper system (Life Technologies) with biotinylated probes 5'-ATGCACA CCGTGCGTACGTCGC-3' and 5'-GACCGTGACCAACTTCTACATCGCA-3'. *hOT7T175* was stably expressed in CHO/*dhfr*-cells using pAKKO-1.11H vector¹⁵ and in B16-BL6 melanomas using pME18S/neo vector.

Isolation of metastin

The acid extract⁷ from human placenta (740 g, Katayama Chemical) was fractionated using a TSKgel ODS80-TM column (Tosoh, 21.5 × 300 mm)⁷ after the protein precipitation, ether extraction and preparative C18 (Waters, 200 ml) chromatography⁷. Active fractions screened by monitoring [Ca²⁺]_i response (FLIPR assays¹⁶) were subjected to SP-Sephadex C-25 (0.8 × 3 cm, elution: 2 M pyridine acetate) and Sephadex G50 (1.4 × 80 cm) chromatography⁷. The active peptide was further purified with TSKgel CM-2SW (4.6 × 250 mm, 10–500 mM HCOONH₄ gradient in 10% CH₃CN for 60 min at 1 ml min⁻¹), TSKgel Super-Phenyl (4.6 × 100 mm, 21–27% CH₃CN gradient in 0.1% TFA for 60 min at 1 ml min⁻¹), and TSKgel Super-ODS (4.6 × 100 mm, 27–39% CH₃CN gradient in 0.1% HFBA for 60 min at 1 ml min⁻¹) columns.

Receptor binding assay

Assays were done as described⁷. Membrane fraction was prepared using homogenizing buffer (10 mM NaHCO₃, 2 mM EGTA, 0.2 mM MgCl₂, protease inhibitors, pH 7.4) and stored in 50% glycerol–50% homogenizing buffer at –20 °C. Metastin (40–54) was labelled with ¹²⁵I-Na using lactoperoxidase, and purified to a carrier-free single peak.

Cell migration and Matrigel invasion assays

A polyvinylpyrrolidone-free polycarbonate framed filter (5-μm pores coated with 10 μg ml⁻¹ bovine fibronectin for CHO, 8-μm pores coated with 100 μg ml⁻¹ bovine gelatin for B16-BL6 melanomas) was set in a microchemotaxis chamber (Neuro Probe). Cells (2 × 10⁵ cells) and peptide were added to the upper chamber and incubated at 37 °C for 6 h (or 7 h for B16-BL6 melanomas) to allow migration to the lower chamber, which contained 10% FCS/DMEM–0.5% BSA (or 5 mg ml⁻¹ bovine fibronectin/RPMI 1640–0.5% BSA for B16-BL6 melanomas). After removing cells from the upper surface of the filter, cells on the lower surface were fixed, stained with Diff-Quick (International Reagent) and counted by measuring A₅₇₀.

Cells and peptide (5 × 10⁵ cells per 200 μl DMEM–0.5% BSA) were added to a Matrigel (10 μg per well, Collaborative Res.)-coated Transwell (5-μm pores, Costar) and incubated at 37 °C for 6 h versus a lower chamber containing 10% FCS/DMEM–0.5% BSA. After removing the Matrigel and cells from the upper surface of the membrane, cells on the lower surface were fixed, stained with haematoxylin and eosin, and counted in 0.25-mm² fields.

Visualization of focal adhesions and stress fibres

Melanomas (1.5 × 10⁵ cells) were plated in collagen IV (Gibco-BRL)-coated chamber slides (Falcon), incubated in RPMI 1640 suspension medium at 37 °C for 2 h, challenged with metastin at 37 °C and then fixed. After washing with 0.3% Triton X-100 PBS (TPBS) and blocking with 10% goat serum/TPBS (STPBS), cells were incubated with anti-vinculin antibody (Sigma) at 4 °C overnight, washed with TPBS and stained with FITC-conjugated anti-mouse IgG (Jackson) at room temperature for 2 h. Cells were further washed with TPBS, stained with rhodamine phalloidin (Molecular Probe) at room temperature for 20 min and washed with TPBS. Cells were mounted with VECTASHIELD and examined under a Leica TCS II confocal microscope.

Phosphorylation of FAK and paxillin

Melanomas (4 × 10⁶ cells in RPMI 1640) were plated onto collagen-IV-coated 6-cm dishes, incubated at 37 °C for 2 h, challenged with metastin at 37 °C and lysed with 1 ml of lysis buffer (50 mM Tris, 150 mM NaCl, 1 mM EGTA, 2 mM Na₃VO₄, 50 mM NaF, 1% NP-40, 4 mM Na₂P₂O₇ and protease inhibitors, pH 7.4). Each 0.45-ml lysate was precipitated with anti-FAK (Upstate) or anti-paxillin (Zymed Laboratory) antibody using protein G-Sepharose. One half of the precipitate was western blotted using ECL phosphorylation detection (Amersham Pharmacia), and the other half was western blotted using anti-FAK or anti-paxillin antibody.

Spontaneous and experimental tumour metastasis assays

Melanomas (3 × 10⁵ cells per 20 μl) were injected into the forefoot fat pad of female C57 BL/6 6-week-old mice on d0. On d18, an Alzet pump (duration, 14 days, pumping rate, 0.25 μl h⁻¹, Model 1002, Alza) containing 1 mM metastin (100 μl in distilled water) was implanted subcutaneously. On d21, tumours were surgically removed. On d35, the mice were killed and the lungs were fixed with Bouin's fixative. Tumour foci were counted macroscopically. For experimental metastasis assays, an Alzet pump was implanted three days before cells (5 × 10⁴ cells per 100 μl) were injected through the tail vein on d0. The mice were killed on d14.

Quantitative RT-PCR analyses

Analyses were made using a Prism 7700 sequence Detector (PE Biosystems). Primers and fluorescence-labelled probes were: 5'-ACTCACTGGTTCTTGGCAGC-3', 5'-ACCTTTCTAATGGCTCCCA-3' and 5'-Fam-ACGTGCTTCTCTGTGCCACCC ACT-Tamra-3' for *KiSS-1*, or 5'-CGACTTCATGTGCAAGTTCGTC-3', 5'-CACACTCA TGGCGGTCAGAG-3' and 5'-Fam-ACTACATCCAGCAGGTCTCGGTGCAGG-Tamra-3' for *hOT7T175*.

Received 18 December 2000; accepted 3 April 2001.

- Lee, J. H. *et al.* *KiSS-1*, a novel human malignant melanoma metastasis-suppressor gene. *J. Natl Cancer Inst.* **88**, 1731–1737 (1996). Erratum in *J. Natl Cancer Inst.* **89**, 1549 (1997).
- Lee, J. H. & Welch, D. R. Identification of highly expressed genes in metastasis-suppressed chromosome 6/human malignant melanoma hybrid cells using subtractive hybridization and differential display. *Int. J. Cancer* **71**, 1035–1044 (1997).
- Lee, J. H. & Welch, D. R. Suppression of metastasis in human breast carcinoma MDA-MB-435 cells after transfection with the metastasis suppressor gene, *KiSS-1*. *Cancer Res.* **57**, 2384–2387 (1997).
- West, A., Vojta, P. L., Welch, D. R. & Weissman, B. E. Chromosome localization and genomic structure of the *KiSS-1* metastasis suppressor gene (*KISS1*). *Genomics* **54**, 145–148 (1998).
- Lee, D. K. *et al.* Discovery of a receptor related to the galanin receptors. *FEBS Lett.* **446**, 103–107 (1999).
- Bloomquist, B. T. *et al.* Cloning and expression of the human galanin receptor GalR2. *Biochem. Biophys. Res. Commun.* **243**, 474–479 (1998).
- Ohtaki, T. *et al.* Isolation and cDNA cloning of a novel galanin-like peptide (GALP) from porcine hypothalamus. *J. Biol. Chem.* **274**, 37041–37045 (1999).

8. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
9. Rouille, Y. *et al.* Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: the subtilisin-like proprotein convertases. *Front. Neuroendocrinol.* **16**, 322–361 (1995).
10. Nishimura, O. *et al.* Synthesis of new peptides with prolactin-releasing activity by a combination of recombinant DNA technology and a cysteine-specific cyanylation reaction. *Chem. Pharm. Bull. (Tokyo)* **46**, 1490–1492 (1998).
11. Turner, C. E. Paxillin and focal adhesion signalling. *Nature Cell Biol.* **2**, E231–E236 (2000).
12. Sadahira, Y., Ruan, F., Hakomori, S. & Igarashi, Y. Sphingosine 1-phosphate, a specific endogenous signaling molecule controlling cell motility and tumor cell invasiveness. *Proc. Natl Acad. Sci. USA* **89**, 9686–9690 (1992).
13. Yamamura, S., Hakomori, S., Wada, A. & Igarashi, Y. Sphingosine-1-phosphate inhibits haptotactic motility by overproduction of focal adhesion sites in B16 melanoma cells through EDG-induced activation of Rho. *Ann. NY Acad. Sci.* **905**, 301–307 (2000).
14. Bischof, P., Meisser, A. & Campana, A. Paracrine and autocrine regulators of trophoblast invasion. *Placenta* **21**, S55–S60 (2000).
15. Masuda, Y. *et al.* Receptor binding and antagonist properties of a novel endothelin receptor antagonist, TAK-044 [cyclo[D- α -aspartyl-3-[(4-phenylpiperazin-1-yl)carbonyl]-L-alanyl-L-aspartyl-D-2-(2-thienyl) glycy]-L-leucyl-D-tryptophyl]disodium salt], in human endothelin A and endothelin B receptors. *J. Pharmacol. Exp. Ther.* **279**, 675–685 (1996).
16. Fujii, R. *et al.* Identification of neuromedin U as the cognate ligand of the orphan G protein-coupled receptor FM-3. *J. Biol. Chem.* **275**, 21068–21074 (2000).
17. Linden, J. Calculating the dissociation constant of an unlabeled compound from the concentration required to displace radiolabel binding by 50%. *J. Cyclic Nucleotide Res.* **8**, 163–172 (1982).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Sasada, M. Mori, T. Henta, M. Harada, K. Oda and Y. Horikoshi for technical advice and assistance.

Correspondence and requests for materials should be addressed to T.O. (e-mail: Ohtaki_Tetsuya@takeda.co.jp). The nucleotide sequence data reported here will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB051065 (*hOT7175*) and AB051066 (*rOT7175*).

erratum

Self-assembly of mesoscopically ordered chromatic polydiacetylene/silica nanocomposites

Yunfeng Lu, Yi Yang, Alan Sellinger, Mengcheng Lu, Jinman Huang, Hongyou Fan, Raid Haddad, Gabriel Lopez, Alan R. Burns, Darryl Y. Sasaki, John Shelnutt & C. Jeffrey Brinker

Nature **410**, 913–917 (2001)

A symbol was omitted from the author list. Yi Yang should have been denoted as contributing equally to this work with Yunfeng Lu. □

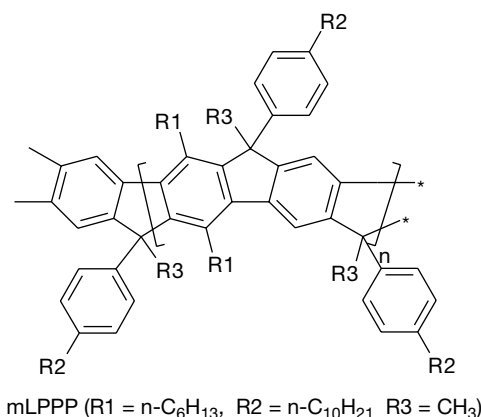
corrections

Formation cross-sections of singlet and triplet excitons in π -conjugated polymers

M. Wohlgenannt, Kunj Tandon, S. Mazumdar, S. Ramasesha & Z. V. Vardeny

Nature **409**, 494–497 (2001)

The molecular structure of the π -conjugated polymer mLPPP shown as an inset to Fig. 1a is incorrect; the correct structure is shown below. Furthermore, we regret not to have acknowledged U. Scherf and K. Muellen, who synthesized the mLPPP polymer. □



Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses

Carin C. Stamper, Yan Zhang, James F. Tobin, David V. Erbe, Shinji Ikemizu, Simon J. Davis, Mark L. Stahl, Jasbir Seehra, William S. Somers & Lidia Mosyak

Nature **410**, 608–611 (2001)

The Protein Data Bank accession number given was incorrect. The correct accession number is 1I8L. □

Michaelis–Menten and beyond

The 'textbook' enzyme kinetics plots, still part of the kinetics art.



TriChromRanger, in 1 × 48 well format.

TriChromRanger

Pierce

www.piercenet.com

Ready-to-go molecular weight markers

TriChromRanger is a mixture of seven pre-stained proteins in the range 16.5–210K. The second highest molecular weight component (phosphorylase B) and the lowest (lysozyme) are covalently modified with red and purple dye, respectively, to act as reference bands within the gel. The five remaining proteins have been covalently modified with a blue dye. The inclusion of two differently stained bands in the marker allows easy marker protein recognition on Western blots or SDS-PAGE gels, should the dye front run off the bottom of the gel.

Enzyme kinetics module

SPSS

www.spss.com

Custom graphing for SigmaPlot

This new module extends SigmaPlot's curve-fitting and graphing capabilities. It features a data entry wizard for transferring data into a SigmaPlot worksheet, including material cut-and-pasted from other software packages or entered manually. Replicate velocities may be entered and, for inhibition studies, substrate or inhibitor concentrations can be varied for each series. There is a built-in library of 11 equation groups and 35 model types. Graph types include Michaelis–Menten, Lineweaver–Burke, Eadie–Hofstee and Scatchard.

QuantiPro

Sigma-Aldrich

www.sigma-aldrich.com

High-sensitivity protein assay

QuantiPro is designed to assay protein concentrations down to $0.05 \mu\text{g ml}^{-1}$ with low sensitivity to interfering agents, a stable colour reaction and resistance to detergents. The bicinchoninic acid (BCA) assay relies on the protein-depen-

dent reduction of copper under alkaline conditions. BCA forms a purple-blue complex with reduced copper in this alkaline environment. Applications include microtitre plate assay and determination of functional groups on solid supports and in solutions.

CoreHTS

PanVera

www.panvera.com

Tyrosine kinase assay with a competitive edge

Fluorescence polarization is the detection method used in the 'Tyrosine Kinase Assay Kit, Green', which is designed for high-throughput drug discovery screening of novel tyrosine kinase inhibitors. The assay is based on competition between a fluorescein-labelled phosphopeptide tracer and the non-fluorescent phosphopeptides generated in the enzyme reaction, binding to an anti-phosphotyrosine antibody. The antibody and the fluorescent phosphopeptide tracer are supplied as separate components in the kit, which is available in 100, 1,000 and bulk assay sizes.

Substrate profiling libraries

ChemRx

www.chemrx.com

New screening tools for protease and GPCR

The ChemRx tripeptidyl aminomethyl coumarin (AMC) library can be used in protease screening for novel substrates for known and new protease targets and to gain information on P1-P3 binding specificity for inhibitor development. And ChemRx's G-protein coupled receptor (GPCR) targeted library is designed for use as a screening tool in the identification of lead compounds.

DuoLuX

Vector Laboratories

www.vectorlabs.com

Alkaline phosphatase and peroxidase substrates

Both these substrates can be used in 'blot assays' or in ELISA applications. Sensitivity in western

blot chemiluminescent detection is approximately 1 pg of protein using either the phosphatase or peroxidase. For nucleic acid detection, the alkaline phosphatase substrate is slightly more sensitive than that for peroxidase, enabling visualization of 100 fg of target nucleic acid. Peroxidase substrate is supplied in two 100-ml bottles, mixed to yield 200 ml of substrate; alkaline phosphatase substrate is ready to use in a 100 ml package.

Freezer racks

Nalgene

www.nalgenunc.com

Cryogenic storage for multiwell plates

Two new designs are available in Nalgene's range of freezer storage racks. A 16-shelf horizontal microwell plate rack in a 4 × 4 array holds 48 standard 96-well plates or 16 deep well plates, while a larger 24-shelf unit will accommodate 72 standard 96-well plates or 24 deep well plates. The stainless steel racks have a retainer clip to hold plates in place, top and side handles and a label holder.

These notes are compiled in the Nature office from information provided by the manufacturers.



Colour-code your freezer storage.

ADVERTISEMENTS

Polyclonal Antibodies & Custom Peptides

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen[™]

Invitrogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at
www.resgen.com

Ubiquitinylation assays?

**Ub-activating enzyme E1
Ub-conjugating enzymes
(E2s)**

**Fractions I & II
Ub conjugating kit**

www.proteasome.com

AFFINITI Research Products Ltd.
Mamhead Castle, Exeter, EX6 8HD, U.K.
T. +44(0) 1626 891010 – F +44(0) 1626 891090
– Email info@proteasome.com –

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: classified@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

European Sales Consultant:

Jonathan Bull (4973)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Sille Opstrup (4994)

Production Manager:

Billie Franklin
To send films and materials use
London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: classified@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Spain/ Portugal:

Amanda Wren
Tel + 34 91 3913835
Fax + 34 91 319 0091
e-mail: a.wren@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: classified@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

Shin-Mitsuke Building (4F),
3-6 Ichigaya Tamachi Shinjuku-Ku,
Tokyo 162
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejpn.com
Japan Manager: Kate Cowan

A tale of two energy policies

The approaches of Europe and the United States to renewable energy research could not be more polarized (see Careers and Recruitment features on pages 4–8). Of course, the political realities reflect the fossil-fuel situations and popular opinion of each. The United States has more coal and gas available if it has the will to tap into sensitive areas, such as the Arctic, and the gall to toss aside the Kyoto Protocol's guidelines on CO₂ emissions. Europe has neither an abundance of fossil fuels nor does the populace have the will to ignore Kyoto.

But political realities and environmental interests aside, what is the practical upshot for jobs — especially in research? The proposed European approach, to fund more research in renewables, as well as implement policies that encourage renewable use could, if successful, create more jobs in science. This, in turn, could help create new industries — perhaps the energy equivalent of the information technology or biotech booms of recent years.

The US approach — to continue to emphasize traditional energy sources over renewables — is sure to hurt US science in the short term. But it could have the ironic effect of hurting US business — the very interest President George W. Bush's policies seem crafted to protect — in the long term.

As David J. Slawson, president of Stirling Energy Systems, a Phoenix-based renewable energy technologies company says, the choice between the president's policy and Europe's direction is obvious. "If the president's budget were to slash and burn the renewables budget and everything we've worked for, we would put all our efforts and expertise into Europe, and walk away from the United States."

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

US emphasis on conventional fuels could undermine renewables research p4

European emphasis on alternative energy could supply jobs and power p7

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Renewable-energy funds threatened

The latest budget proposals from the Bush administration present a mixed bag for US energy R&D, says Steve Bunk.

Controversial plans to drill for oil in Alaska's Arctic National Wildlife Refuge will upset more than just the caribou grazing on this land.

[Far right] Special branch: energy crops hold the key to the future of biomass energy.

NREL: where proposed cuts threaten up to 400 jobs in energy research.

March blackouts in California were followed by an April response from President George W. Bush that was ominous for those engaged in the research and development of power from renewable resources. Bush's budget proposed that Congress approve funding cuts in renewable energy of more than 36% for fiscal year 2002 compared to this year. Solar research would suffer a 54% reduction, to \$42.9 million, and wind 48%, to \$20.5 million.

Lobbying might restore funding to this year's levels when the final budget is approved, especially given the national energy policy released in May. It ties extra funding for renewable energy to royalties from proposed escalations in exploration and development of conventional resources on federal lands, such as the controversial plan to drill for oil in Alaska's Arctic National Wildlife Refuge. But the threat of cuts — along with such moves as rejecting the Kyoto treaty on global warming, reversing a campaign pledge to regulate carbon dioxide emissions and proposing a \$2 billion outlay over 10 years to develop 'clean coal' technology — bodes ill for US renewable-energy research.

Opponents of the proposed cuts say that, if enacted, they would reduce the 840-strong staff at the federal government's National Renewable Energy Laboratory (NREL) in Golden, Colorado, by up to a

half. NREL employs geneticists for bioenergy development, materials scientists for solar technologies, aerodynamics experts for wind power, geologists for geothermal power, chemical engineers for superconductivity, biophysicists for hydrogen research and mechanical engineers for energy-efficient building design. "NREL requires as many disciplines in that

relatively little lab as I've seen in any one place," says director Richard Trully. But much of their research suffered deep cuts in the proposed budget. Moreover, the renewables industries have traditionally been slow job markets, because of perennial losses. As these varied technologies tread their individual paths toward commercial viability, how will the chilly political climate affect research?

BIOMASS APPEAL

The most power produced by renewable energy sources in North America comes from biomass, such as forest industry and agricultural residues. Electricity from biomass comes primarily from small producers such as pulp- and paper-mills that burn their waste to drive steam turbines. But a thermochemical conversion technology developed in the 1980s at the Battelle Memorial Institute in Columbus, Ohio, can circumvent the steam cycle by converting biomass to gas burned directly in turbines. Future Energy Resources Corp. (FERCO) of Atlanta, Georgia, was formed in 1992 to buy the Battelle technology, which has received about \$27 million in US Department of Energy developmental funding.

"Biomass has its place but very few new plants are being built. Right now, it's probably overshadowed by other renewable energies," says FERCO business development director Sim Weeks. "That's what we'd like to change."

Nevertheless, bioenergy research is unharmed by the president's budget, with proposed funding just below this year's level. A National



Bioenergy Center was formed last November at NREL. Growing 'energy crops' rather than using waste products is where biomass will exert its main influence on the energy mix, says Janet Cushman, product manager of the biofuels feedstock development programme at Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee. However, she admits, "there's a long way to go toward developing an infrastructure for non-food or non-forest products".

For these fast-growing trees and grasses, skills in plant-breeding, crop selection and development, and production management are necessary. "Some of the most interesting work going on is trying to decide among alternative applications in new bioprocessing facilities," Cushman says. Chemical engineering, microbial biology, genetics, bioprocessing and virtually all the plant and agricultural sciences are involved, says Cushman, an ecologist. Handling and processing of biomass, including feeding it into the system, create challenges that have induced ORNL to hire its first agricultural engineer.

SOLAR UNCERTAINTIES

"Access to the federal labs is extremely important, because of 30 years of a brain trust there," says David Slawson, president of Stirling Energy Systems, a renewable energy technologies company in Phoenix, Arizona. "If the president's budget were to slash and burn the renewables budget and everything we've worked for, we would put all our efforts and expertise into Europe, and walk away from the United States."

The company's main product is a solar-dish system that includes a collection of mirrors that concentrate sunlight to provide heat, expanding hydrogen to rotate an electrical generator. Slawson says many of his key workers formerly were retired mechanical and electrical engineers who helped to initiate the technology in the mid-1970s, and who now prefer to work part-time. The company also trains young engineers.

Although such concentrating solar power (CSP) systems differ from the solar cells of photovoltaic devices, Slawson believes that many politicians don't understand CSP. "They seem to think that solar only has to do with photovoltaics and water heaters that are expensive and haven't had a good record." CSP electricity costs around 10 cents per kilowatt-hour (kWh), whereas photovoltaics is about 20 cents per

kWh. In Spain and Italy, subsidies provide 19 cents per kWh for solar energy above a base rate of 6 cents per kWh, Slawson says. Stirling Energy Systems projects that its dish technology will be price-competitive by 2003, and that generating capacity will reach 750 MW by 2005.

The 86% federal funding reduction proposed for CSP "could be a devastating blow", says Terry Peterson, manager for solar power at the Electric Power Research Institute, a Palo Alto, California-based collaborative science and technology organization for the power industry. "The budget is a disaster." But he adds that, given how close these technologies are to commercial cost-effectiveness, it is not so critical that the federal R&D budget be optimal.

Nevertheless, Slawson and other solar industry leaders recently had high-level talks in Washington, which gave him hope that federal funding will be restored. CSP can have either small or massive applications, he says. "If everything went well, we would be able to build the equivalent of the Hoover Dam (which produces 2.2 GW) every two years in mass production."

WIND AND STEAM

The most mature and fastest-growing of the renewable technologies in North America is wind power. "We're going to have a surge in investments this year, on the order of a 70% increase," says Christine Real de Azua, communications director of the Washington DC-based American Wind Energy Association.

Windmills currently supply about 2 GW of power, which is less than 1% of total needs, but another 1 GW will go online during 2001. Enron Corp. of Houston,



Taking the heat: geothermal (top) and solar (bottom) power research will be hit hard by the proposed budget cuts.



Hydrogen power

Hydrogen, the most plentiful element in the Universe, is often nominated as the ideal fuel, especially as its only by-product is water. But widespread use remains decades away, and skills required to achieve that end will probably continue to expand as research progresses.

The future of hydrogen power lies with interdisciplinary collaborators, says Anastasios

Melis, a professor of plant and microbial biology at the University of California, Berkeley. "It's very hard to get one person who is an expert in this many diverse technologies." Last year, Melis's lab announced the discovery of a molecular switch that can regulate the generation of hydrogen as a metabolic by-product of the green microalga *Chlamydomonas reinhardtii*.

The microalgae are engineered to enhance their hydrogen production, a mutagenesis that demands skills in the overlapping fields of microbiology and molecular genetics. The lab is now scaling up the process from one-litre flasks to a 1,000-litre photobioreactor for which Melis is bringing in mechanical and chemical engineers. The technique, one of several

possibilities for sustained production of the gas, has attracted interest from potential investors.

Melis expects that the earliest use of hydrogen will be for the generation of electricity through fuel cells or hydrogen turbines. "And as we move toward commercialization of the process, it appears that additional skills will be needed."

S. B.

Texas, will complete construction late this year in Texas of a 135-MW wind facility, enough to power 50,000 homes. Such developments bode well for the range of wind-power jobs, including electronics, hydraulics, design of drive trains and airfoils, and information technology. Enron is currently looking for people with expertise in electromagnetics, cryogenics, rotor dynamics, composites fabrication, stress analysis and anemometry.

Bob Gates, Enron's vice-president of business development, foresees major wind farms in the Great Plains states east of the Rocky Mountains. He predicts growth in wind generation to 10–15% of total national production in coming years. Natural gas currently costs 5–6 cents per kWh of electricity production, whereas wind costs 4–8 cents per kWh. Gates declares that much job satisfaction in the wind industry derives from helping to reduce environmental damage.

Geothermal research was also hit hard by the Bush budget. Funding was eliminated for the Clinton administration's plan to expand geothermal power in the west, and the total budget fell by 49%. Steam from such resources is highly site-dependent. The largest source, near Northern California's Napa Valley, currently produces 1 GW, enough electricity for about one million homes. Most of that power is provided by Calpine of San Jose, California, which is adding about 150 MW to its capacity by piping treated waste-water from nearby communities into the site. However, Calpine's primary focus is on developing natural-gas resources, the major growth area in US energy.

SUPERCONDUCTIVITY

Energy-efficiency technologies, which are related to the renewables field, were also targeted for cuts. Among the most promising of these is superconductivity, or the ability of certain materials to conduct electricity with essentially no resistive losses. The president's plan called for a reduction in superconductivity research from this year's \$32 million to \$19 million. However, John Howe, vice-president of electric industry affairs at American Superconductor Corp. in Westborough, Massachusetts, saw a written request from the Department of Energy that superconductivity research be brought back to current funding levels.

Over the past decade, American Superconductor has developed a ceramics-based wire that can carry 100 times more power than a conventional copper conductor of the same dimensions. In early 2002, the company will complete construction in central Massachusetts of the world's first commercial-scale manufacturing plant for high-temperature superconducting (HTS) wires. The company is now installing its first test in a utility grid in downtown Detroit.

"A lot of the R&D work, in terms of basic materials, is already under our belt," Howe says. "What's important now is how to integrate these materials into commercial-scale devices,

Buildings

When he talks buildings, Paul Torcellini gets on a missionary roll. "For the most part, buildings in this country are a non-engineered product," says the senior engineer for the building energy research team at the National Renewable Energy Laboratory in Golden, Colorado. "If you had to buy a car the same way as you buy a building, the car would never get out of the garage."

Buildings, he says, consume 36% of US energy for their operation and account for roughly the same percentage of pollution. When people think of pollution and environmental damage, they usually consider transportation and industry, he says. "But the real story is, 68% of the electricity consumed in this country goes into buildings, and most of that electricity is coal-fired."

Much of Torcellini's time is spent writing and speaking about high-performance buildings, which to him

means climate-sensitive design that integrates the best of the old and the new. Some old buildings, without the benefits of electricity, were well-engineered. But most new buildings, exemplified by strip malls, are concerned primarily with low construction costs.

Using design ideas that go back a century or more, access to natural light and breezes can save energy and reduce cooling needs. But to engineer a building nowadays requires the collaboration of architects, structural and mechanical engineers, landscapers, interior designers, the construction crew, management and maintenance people to use the latest monitoring technologies properly.

Torcellini sees recent electricity blackouts as a two-sided issue. "Why do we need more supply? Because we consume too much. So you've got to do research on both sides, to balance the equation."

S. B.

and how they'll operate under actual field conditions." The company, which has about 400 employees, seeks expertise in chemical engineering, materials science, cryogenics, electromagnetics, mechanical and electrical engineering, experimental physics, computer sciences and practical design engineering.

Steve Millett, a futurist at Battelle Memorial Institute, believes renewable resources have been largely forgotten in the US energy mix. "I think renewables is almost virgin territory," he says. "That may mean it will be a dot.com of the future."

Steve Bunk is a science writer in Albuquerque, New Mexico.

Web links

Energy Efficiency and Renewable Energy Network

♦ <http://www.eren.doe.gov>

National Renewable Energy Laboratory ♦ <http://www.nrel.gov>

Solar Energy Industries Association ♦ <http://www.seia.org>

American Bioenergy Association ♦ <http://www.biomass.org>

American Wind Energy Association ♦ <http://www.awea.org>

National Hydrogen Association ♦ <http://www.ttcorp.com/nha>

Full power? Energy-efficiency technologies such as superconductivity have also been threatened with cuts.

JIM YOST PHOTOGRAPHY



Alternatives energize Europe

The commitment to invest in renewable energy sources may provide jobs as well as power, says Helen Gavaghan.



HAGESHOLM WINDPOWER

It is a truism in science that the underlying assumptions affect predictions. So it is hardly surprising that studies of current and potential job creation associated with a shift to renewable energies should vary widely depending upon interpretations. Yet one thing seems certain: more rather than fewer jobs will be associated with renewable energies.

An influential policy paper that the European Union (EU) published in 1997 estimated there could be up to 900,000 jobs created across Europe in renewable energies by 2010. "This does not sound unrealistic to me," says Leon Freris, a professor at the University of Loughborough in Britain and joint founder of the university's Centre for Renewable Energy Systems Technology (CREST). "In 2001 the figures for market penetration of wind power alone had already been exceeded by 50%, so the numbers might even be an underestimate," he says.

Although many of the new jobs will be in construction or installation of power plants, there is a need for people to perform a wide range of tasks. "Remember that renewable energies sit amid a wide range of institutions from venture capital for start-up companies to banking, planning, regulation and power distribution," says Catherine Mitchell, currently seconded from Warwick University to the Performance and Innovation Unit of the UK Cabinet Office. Her task is to advise the prime minister, Tony Blair, on how an additional £100 million (US\$142 million) of government funding announced in March this year (additional to some £150 million already committed to wind farms, biomass, and research and development) should be spent during the next three years.

In particular, argues Mitchell, there is a need for people to think and research how renewable energy sources can be connected to the power grid. Sven Svendsen, a professor of building energy technology at the Danish Technological Institute agrees. Distribution systems are designed to take specific amounts of

electricity from large gigawatt power stations, he says, and not comparatively small, unpredictable amounts from wind farms or solar collectors, where the output depends on weather.

Current distribution systems can take perhaps 25% of electricity from renewable sources without major modifications, says Freris, formerly head of power systems at Imperial College, London, but as the portion of electricity from renewables grows there is a need for research into how they can be integrated into a secure supply. For example, says Ole Langniss, a specialist in renewable-energy policy from the German Aerospace Research Centre (DLR), the engineering group ABB (Asea Brown Boveri) is researching how wind turbines can be connected directly to the high-voltage distribution system and not via a transformer.

POLICY DRIVERS MAKE RENEWABLES INEVITABLE

Although there are problems to solve and the solution will depend on the mixture of renewable energy sources adopted, the powerful economic, environmental and national security factors influencing European energy policy mean that a trend towards renewables is inevitable says Svendsen.

EU policy-makers are deeply conscious that the EU imports more than half of its energy needs and, says a European Commission green paper on energy supply security published last November, if nothing is done, that figure will rise to 70% in 20 to 30 years. At the same time, EU policy-makers are discussing ways to make sure that greenhouse gas emissions meet the targets set at Kyoto in 1997, while ensuring economic growth continues.

Powering ahead: Denmark is aiming to produce 30% of its electricity from wind turbines.

Building energy: innovative use of renewable resources may help the EU redress its energy import imbalance.



COLIN CUTHBERT/SPL

To that end, many national and pan-European initiatives focus on increasing the role of renewables. Two specific targets have been adopted informally: to increase the contribution that renewables make to energy consumption in general from 6% in 1997 to 12% in 2010; and to increase the percentage of electricity generated from renewables from 14 to 22% in the same period. Unless there is also a parallel reduction in energy demand from consumers, the economic growth of the past few years would make the renewable targets hard to meet.

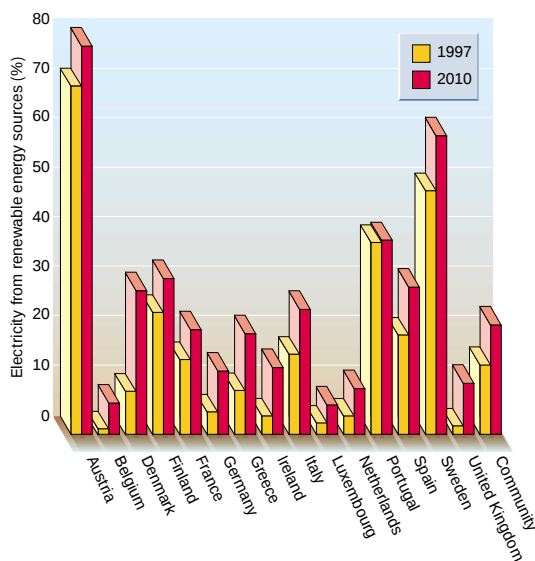
Faced by these realities, European countries are pressing ahead with their own national plans for energy efficiency and conservation, and to increase the proportion of electricity generated from renewables (see graph below).

NEW COURSES AND ACADEMIC POSITIONS

One indication of the surge of interest in renewable energies is an advertisement for one professor, a senior lecturer and a lecturer at CREST. Such largess by university administrators is rare indeed. CREST also has plans for a new undergraduate course dedicated to renewable energies to start in 2002.

Part of the thinking behind the undergraduate course is to attract people to study power systems for renewable energies, says Freris. Loughborough already offers a masters course in renewable energy. This is a technical course focusing on the mature renewable technologies of wind, photovoltaics, biomass and water power. It suits engineering or physics graduates, says Freris, as well as people in industry wishing to develop their knowledge. For the future, CREST hopes to see a European-wide masters course in renewable energies established with credentials that are recognized continent-wide.

One of the youngest masters courses is in Sweden at Högskolan Dalarna University. Here students focus on solar energy. Those students with an undergraduate degree in mechanical engineering can get a masters, others receive a diploma in solar energy technology, although the university is in the process of developing its masters programme further. The course is taught in English, and students can gain research experience by working part of the time with the university's Solar



EU national targets for the contribution of electricity produced from renewable energy sources compared with gross electricity consumption by 2010.

Mature renewable options

Hydropower is clearly the best established alternative energy source, and there is unlikely to be much further development of this in Europe, says Leon Freris, a professor at the Centre for Renewable Energy and Systems Technology at Loughborough University, UK. Of the 'new' renewable sources that are close to rivalling the economics of fossil fuels, wind is the best developed and is increasing its market share rapidly.

Denmark, Germany and Spain have made most progress with wind power, and Denmark and Germany have the best-developed industrial base for wind turbines. By 2001, Denmark was half way to its target of

generating 30% of electricity from wind, largely because of its commitment to wind power and its offshore windfarms. And last month Britain announced that preliminary permission has been granted to 18 wind-farm developers for offshore wind sites.

Britain has also introduced a levy on electricity from oil and gas and is passing legislation obliging electricity companies to buy at least 10% of their supplies from renewable providers. The story is the same around Europe. Governments are weaving together research, fiscal and market policies that will drive energy production steadily in the direction of energy efficiency and renewable sources. H. G.

Energy Research Centre, says Lars Broman, the professor who founded the course.

The kind of jobs that students from these courses go on to varies. "From Loughborough the majority go in to the renewable energy industry which in the UK means installers, not manufacturers," Freris says. Installers assess wind resources, design wind farms and analyse their connections to the distribution networks. Other graduates become consultants or work in local planning and policy making.

Most jobs are in commerce or industry, but there are openings in academia as CREST's current recruitment shows. Research topics might be the aerodynamics of wind blades or noise abatement, says Ole Langniss of the German Aerospace Research Centre, or they could be in photovoltaic films, or modelling the impact of renewables on distribution centres. One thing seems certain, whether in industry, commerce or academia, renewable energies looks set to be a vibrant field for some decades to come.

Helen Gavaghan is a freelance science writer based in West Yorkshire.

Web links

Centre for Renewable Energy Systems Technology

♦ <http://www.lboro.ac.uk/departments/el/research/crest>

European Solar Engineering School ♦ <http://www.eses.org>

A global overview of renewable energy ♦ <http://www.agores.org>

International Energy Agency ♦ <http://www.iea.org>

Department of Trade and Industry ♦ <http://www.dti.gov.uk/renew>

German environment ministry ♦ <http://www.bmu.de/english>